I. P. NATANSON

THEORY OF FUNCTIONS OF A REAL VARIABLE

Volume II

Translated from the Russian by
LEO F. BORON

Theoretical physicists, engineers, senior and graduate students of Lebesgue integrals and related chapters of mathematics will not only find this an invaluable reference book, but will also be delighted by the clarity and elegance of diction and proofs.

AN EXCELLENT TEXT for senior or graduate courses on measure theory, Lebesgue integration, summable functions, functions of finite variation, absolutely continuous functions, singular integrals, point sets in two-dimensional spaces, measurable functions and their integration. These and many other subjects form the contents of these two remarkable volumes which lead in their field.

A number of modifications in the original Russian text make this English version unusually attractive to students who wish to study these subjects either in class or on their own. Theoretical physicists, chemists, engineers, and others who make use of Lebesgue integration in their work will find this an invaluable reference book. Readers will welcome the clarity and elegance of the proofs and the extension to infinite sets of the theorems dealing with the Lebesgue integral.

I. P. NATANSON

THEORY OF FUNCTIONS OF A REAL VARIABLE

(Teoria functsiy veshchestvennoy peremennoy,
Chapters X to XVII and Appendices)

VOLUME II

Translated from the Russian by
LEO F. BORON

FREDERICK UNGAR PUBLISHING CO.
NEW YORK

CONSTABLE & CO LTD
10-12 ORANGE STREET, LONDON, W.C.2

Copyright © 1960 by Frederick Ungar Publishing Co.

PRINTED IN GREAT BRITAIN

Library of Congress Catalog Card Number 54-7420

TRANSLATOR'S NOTE

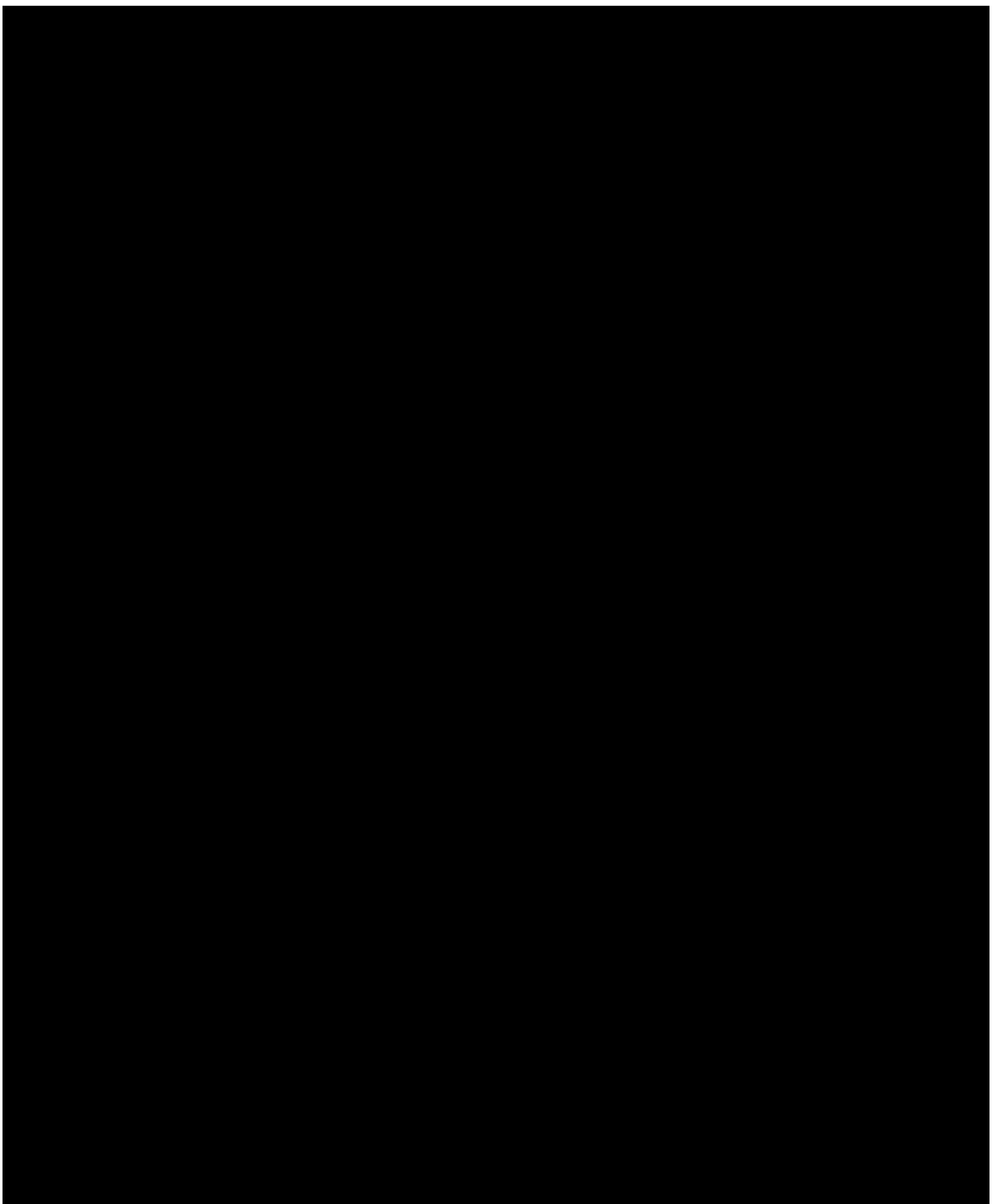
THIS is a translation of Professor Natanson's *Teoriya Funktsii Veshchestvennoi Peremennoi*, Chapters X-XVI, XVIII of the second revised Russian edition, 1957; Chapter XVIII forms the content of Chapter XVII of the present volume. Chapter XVII of the Russian edition was omitted as such because the author follows quite closely (except for Appendix VI) the material in the Boron-Hewitt American edition of Chapters I-IX listed as Editor's Appendices and written by Professor Dr. Edwin Hewitt, dealing with the definition and basic properties of arbitrary measurable sets on the real line, the basic properties of measurable functions defined on sets which are not necessarily bounded, the definition of the Lebesgue integral for functions defined on unbounded sets, square-summable functions defined on arbitrary measurable sets on the real line, functions of finite variation on the infinite line $(-\infty, \infty)$ and the Stieltjes integral of such functions. Appendices I, II, III are the same as in the second Russian edition. Appendix IV, Change of Variable in the Lebesgue Integral, is §5, Chapter IX, 2nd ed. Appendix V, Hausdorff's Theorem, is §5, Chapter XI, 1st ed. Appendix VI, Indefinite Integrals and Absolutely Continuous Set Functions, is §6, Chapter XVII, 2nd ed. Appendix VII, The Role of Russian and Soviet Mathematicians in the Development of the Theory of Functions of a Real Variable, is Chapter XVII, 1st ed.

In this volume, whenever reference is made to Chapters I-IX, it is understood to mean that the reader should consult the first volume of Natanson's *Theory of Functions of a Real Variable*, Chapters I-IX, Frederick Ungar, N.Y., 1955.

The enthusiastic response to the first volume of the American edition prompted the preparation of the remainder of this excellent Russian work for the English-speaking mathematical public. The translator's only hope is that he has reasonably closely approximated the reader's expectations.

LEO F. BORON

Philadelphia, 1960



CONTENTS

CHAPTER X. SINGULAR INTEGRALS. TRIGONOMETRIC SERIES. CONVEX FUNCTIONS **11**

1. Concept of Singular Integral 11
2. Representation of a Function by a Singular Integral at a Given Point 15
3. Application to the Theory of Fourier Series 21
4. Further Properties of Trigonometric and Fourier Series 29
5. Schwarz Derivatives and Convex Functions 36
6. Uniqueness of the Trigonometric Series Expansion of a Function 48

Exercises 57

CHAPTER XI. POINT SETS IN TWO-DIMENSIONAL SPACE **60**

1. Closed Sets 60
2. Open Sets 62
3. Theory of Measure of Plane Sets 64
4. Measurability and Measure as Invariants under Isometries 70
5. The Relation between the Measure of a Plane Set and the Measures of Its Sections 75

CHAPTER XII. MEASURABLE FUNCTIONS OF SEVERAL VARIABLES AND THEIR INTEGRATION **80**

1. Measurable Functions. Extension of Continuous Functions 80
2. The Lebesgue Integral and Its Geometric Interpretation 83
3. Fubini's Theorem 85
4. Interchanging the Order of Integration 90

CHAPTER XIII. SET FUNCTIONS AND THEIR APPLICATIONS IN THE THEORY OF INTEGRATION **94**

1. Absolutely Continuous Set Functions 94
2. The Indefinite Integral and the Differentiation of It 100
3. Generalization of the Preceding Results 103

CHAPTER XIV. TRANSFINITE NUMBERS **107**

1. Ordered Sets. Order Types 107
2. Well-Ordered Sets 112
3. Ordinal Numbers 115
4. Transfinite Induction 118
5. The Second Number Class 119
6. Alephs 122
7. Zermelo's Axiom and Theorem 124

CHAPTER XV. THE BAIRE CLASSIFICATION **128**

1. Baire Classes 128
2. Non-Vacuousness of the Baire Classes 133
3. Functions of the First Class 139
4. Semi-Continuous Functions 149

CHAPTER XVI. CERTAIN GENERALIZATIONS OF THE LEBESGUE INTEGRAL **157**

1. Introduction 157
2. Definition of the Perron Integral 158
3. Fundamental Properties of the Perron Integral 160
4. The Indefinite Perron Integral 163
5. Comparison of the Perron and Lebesgue Integrals 165

6. Abstract Definition of the Integral and Its Generalizations 169

7. The Denjoy Integral in the Restricted Sense (i.e. the Denjoy-Perron Integral) 175

8. H. Hake's Theorem 178

9. The P. S. Aleksandrov-H. Looman Theorem 185

10. The Denjoy Integral in the Wide Sense (i.e. the Denjoy-Khinchin Integral) 189

Exercises 191

CHAPTER XVIII. SOME IDEAS FROM FUNCTIONAL ANALYSIS 193

1. Metric and, in Particular, Normed Linear Spaces 193

2. Compactness 198

3. Conditions for Compactness in Certain Spaces 202

4. Banach's "Fixed-Point Principle" and Some of Its Applications 216

APPENDICES 225

I. Arc Length of a Curve 225

II. Steinhaus's Example 228

III. Certain Supplementary Information about Convex Functions 230

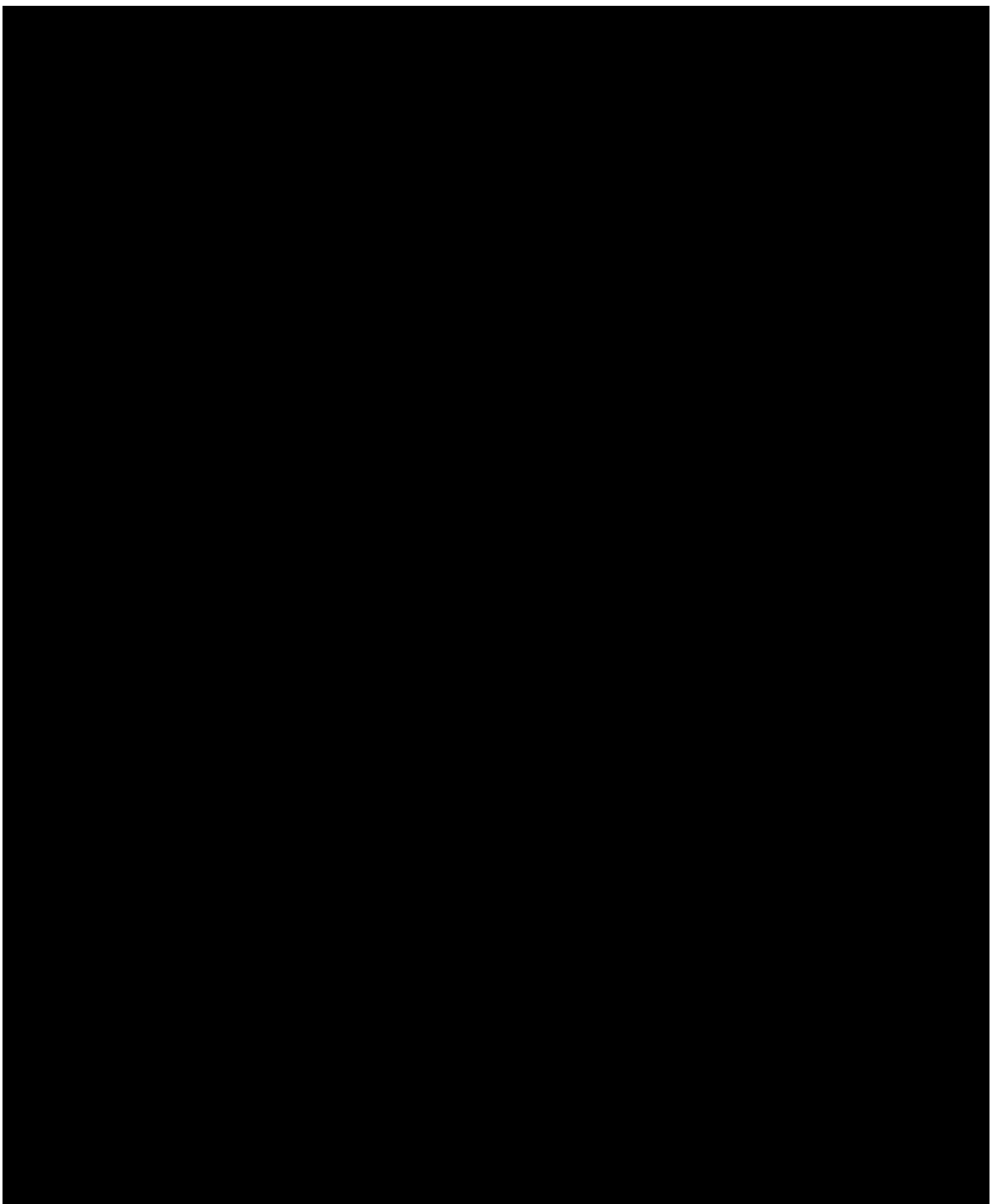
IV. Change of Variable in the Lebesgue Integral 234

V. Hausdorff's Theorem 238

VI. Indefinite Integrals and Absolutely Continuous Set Functions 246

VII. The Role of Russian and Soviet Mathematicians in the Development of the Theory of Functions of a Real Variable 248

INDEX 259



CHAPTER X

SINGULAR INTEGRALS. TRIGONOMETRIC SERIES. CONVEX FUNCTIONS

In this chapter we aim to show the applications of the methods of the theory of functions of a real variable to various problems of analysis.

§ 1. CONCEPT OF SINGULAR INTEGRAL

In order to acquaint the reader with the idea lying at the base of the important concept of singular integral, we shall begin with an example.

Consider the function

$$\Phi_n(t,x)=\frac{1}{\pi}\cdot\frac{n}{1+n^2(t-x)^2}. \tag{1}$$

If n and x are fixed and t varies from 0 to 1, $\Phi_n(t, x)$ is a continuous function of t. Hence we can form the quantity

$$f_n(x)=\frac{n}{\pi}\int_0^1\frac{f(t)\,dt}{1+n^2(t-x)^2} \tag{2}$$

for every summable $f(t)\,(0\leqslant t\leqslant 1)$.

We shall prove that

$$\lim_{n\to\infty} f_n(x)=f(x) \tag{3}$$

at all points $x\,(0<x<1)$ for which the function $f(t)$ is continuous.

To do this, we note first of all that

$$\int_0^1\Phi_n(t,x)\,dt=\frac{n}{\pi}\int_0^1\frac{dt}{1+n^2(t-x)^2}=\frac{1}{\pi}\int_{-nx}^{n(1-x)}\frac{dz}{1+z^2}\to\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{dz}{1+z^2}=1 \tag{4}$$

as $n\to\infty$.

Therefore, to establish (3), it will suffice to show that the difference

$$r_n=f_n(x)-f(x)\int_0^1\Phi_n(t,x)\,dt=\frac{n}{\pi}\int_0^1\frac{f(t)-f(x)}{1+n^2(t-x)^2}\,dt$$

tends to zero as $n\to\infty$.

To this end, having taken an arbitrary $\varepsilon>0$, we find a $\delta>0$ such that $|f(t)-f(x)|<\varepsilon$ for $|t-x|<\delta$. Assuming that $0<x-\delta<x+\delta<1$, we represent r_n in the form

$$r_n = \frac{n}{\pi}\int_0^{x-\delta} \frac{f(t)-f(x)}{1+n^2(t-x)^2}\,dt + \frac{n}{\pi}\int_{x-\delta}^{x+\delta} \frac{f(t)-f(x)}{1+n^2(t-x)^2}\,dt + \frac{n}{\pi}\int_{x+\delta}^{1} \frac{f(t)-f(x)}{1+n^2(t-x)^2}\,dt =$$

$$= A_n + B_n + C_n.$$

The integral B_n is estimated in the following way:

$$|B_n| \leqslant \frac{n}{\pi}\int_{x-\delta}^{x+\delta} \frac{|f(t)-f(x)|}{1+n^2(t-x)^2}\,dt \leqslant \varepsilon \cdot \frac{n}{\pi}\int_{x-\delta}^{x+\delta} \frac{dt}{1+n^2(t-x)^2} \leqslant \frac{\varepsilon}{\pi}\int_{-\infty}^{+\infty} \frac{dz}{1+z^2} = \varepsilon.$$

As for A_n, in this integral we have $|t-x| \geqslant \delta$ so that

$$|A_n| \leqslant \frac{n}{\pi(1+n^2\delta^2)}\int_0^{x-\delta} |f(t)-f(x)|\,dt < \frac{A(\delta)}{n},$$

where $A(\delta)$ does not depend on n. Analogously,

$$|C_n| < \frac{C(\delta)}{n} \text{ and hence } |r_n| < \varepsilon + \frac{A(\delta)+C(\delta)}{n},$$

so that for sufficiently large n, $|r_n| < 2\varepsilon$, i.e. r_n tends to 0 with increasing n, which was to be proved.

It is not difficult to find out really what properties of the function $\Phi_n(t,x)$ assure relation (3). The fact is that for very large values of n, those values of $\Phi_n(t,x)$ which correspond to values of t, which are at an arbitrary nonzero distance from x, are very small, so that the magnitude of integral (2) is determined basically by the values of the function under the integral sign in an immediate neighborhood of the point x. But near the point x, the function $f(t)$ is almost equal to $f(x)$ (since it is continuous for $t = x$). This means that if n is large, then the integral (2) changes little upon replacing $f(t)$ by $f(x)$, i.e. it is almost equal to the integral

$$f(x)\cdot\frac{n}{\pi}\int_0^1 \frac{dt}{1+n^2(t-x)^2}$$

and, by (4), it is then almost equal to $f(x)$.

A function $\Phi_n(t,x)$ possessing these properties is said to be a *kernel*. The exact definition of a kernel follows.

DEFINITION. A *kernel* is a function $\Phi_n(t,x)(n=1,2,\ldots)$ defined in the square $(a \leqslant t \leqslant b, a < x < b)$ and such that

$$\lim_{n\to\infty}\int_\alpha^\beta \Phi_n(t,x)\,dt = 1,$$

provided $a \leqslant \alpha < x < \beta \leqslant b$. It is self-evident that $\Phi_n(t,x)$ is assumed summable with respect to t for every fixed x.

An integral of the form

$$f_n(x) = \int_a^b \Phi_n(t,x) f(t)\,dt,$$

where $\Phi_n(t, x)$ is a kernel, is called a *singular integral.*

The theory of such integrals has numerous applications. The fundamental problem of this theory consists in establishing the connection between limit values of the integral $f_n(x)$ as $n \to \infty$ and the value of the function $f(t)$ at the point x. Since modifying the value of the function $f(t)$ at one point has no effect on the value of $f_n(x)$, it is necessary to require that the value $f(x)$ of the function $f(t)$ at the point x be somehow related to its values at proximate points. The simplest form of such relationship is the continuity of the function $f(t)$ at the point $t = x$. Other forms of this connection are approximate continuity, the requirement that x be a Lebesgue point of the function $f(t)$, and so forth.

We now introduce one of Lebesgue's theorems which will be useful in the sequel.

THEOREM 1 (H. LEBESGUE). *Let $\varphi_1(t), \varphi_2(t), \varphi_3(t), \ldots$ be a sequence of measurable functions defined on $[a, b]$. If there exists a constant K such that*

$$|\varphi_n(t)| < K \tag{5}$$

for all n and t, and if for every c $(a \leqslant c \leqslant b)$, we have

$$\lim_{n\to\infty} \int_a^c \varphi_n(t)\,dt = 0, \tag{6}$$

then the equality

$$\lim_{n\to\infty} \int_a^b f(t)\varphi_n(t)\,dt = 0 \tag{7}$$

is valid for an arbitrary summable function $f(t)$ on $[a, b]$.

Proof. If $[\alpha, \beta]$ is a closed subinterval of $[a, b]$, then (6) implies that

$$\lim_{n\to\infty} \int_a^\beta \varphi_n(t)\,dt = 0. \tag{8}$$

Noting this, we now consider any *continuous* function $f(t)$ and for prescribed $\varepsilon > 0$, we subdivide $[a, b]$ by means of the points $x_0 = a < x_1 < \ldots < x_m = b$ into small subintervals such that the oscillation of $f(t)$ is less than ε in each of them. Then

$$\int_a^b f(t)\varphi_n(t)\,dt = \sum_{k=0}^{m-1} \int_{x_k}^{x_{k+1}} [f(t) - f(x_k)]\varphi_n(t)\,dt + \sum_{k=0}^{m-1} f(x_k) \int_{x_k}^{x_{k+1}} \varphi_n(t)\,dt. \tag{9}$$

But

$$\left| \int_{x_k}^{x_{k+1}} [f(t) - f(x_k)]\varphi_n(t)\,dt \right| \leqslant K\varepsilon(x_{k+1} - x_k),$$

so that the first sum in (9) is not greater than $K\,\varepsilon(b - a)$. The second sum in (9) tends to zero as n increases, in virtue of (8), and hence for $n > n_0$ it will be less than ε. For these n, we have

$$\left| \int_a^b f(t)\varphi_n(t)\,dt \right| < \varepsilon[K(b - a) + 1],$$

so that (7) is proved for a continuous function $f(t)$.

Next, let $f(t)$ be a measurable *bounded* function: $|f(t)| \leqslant M$.

We take $\varepsilon > 0$, and using N. N. Luzin's theorem, find a continuous function $g(t)$ such that

$$mE(f \neq g) < \varepsilon, \quad |g(t)| \leqslant M.$$

Then

$$\int_a^b f(t)\varphi_n(t)\,dt = \int_a^b [f(t)-g(t)]\varphi_n(t)\,dt + \int_a^b g(t)\varphi_n(t)\,dt.$$

But

$$\left|\int_a^b [f(t)-g(t)]\varphi_n(t)\,dt\right| = \left|\int_{E(f\neq g)} [f(t)-g(t)]\varphi_n(t)\,dt\right| < 2KM\varepsilon,$$

where the integral $\int_a^b g\varphi_n\,dt$ tends to zero according to what has already been proved, and, for sufficiently large n, becomes less than ε. Hence, for these n, we have

$$\left|\int_a^b f(t)\varphi_n(t)\,dt\right| < (2KM+1)\varepsilon,$$

which proves (7) for the case of a bounded measurable function.

Finally, let $f(t)$ be an arbitrary summable function.

We take an $\varepsilon > 0$, and using the absolute continuity of the integral, we find a $\delta > 0$ such that for an arbitrary measurable set $e \subset [a, b]$ of measure $me < \delta$ we have

$$\int_e |f(t)|\,dt < \varepsilon.$$

This done, we find a measurable bounded function $g(t)$ (Chapt. IV, §4, Theorem 1) such that

$$mE(f \neq g) < \delta.$$

We may assume that the function $g(t)$ vanishes on the set $E(f \neq g)$. Then

$$\int_a^b f(t)\varphi_n(t)\,dt = \int_a^b [f(t)-g(t)]\varphi_n(t)\,dt + \int_a^b g(t)\varphi_n(t)\,dt.$$

But

$$\left|\int_a^b [f(t)-g(t)]\varphi_n(t)\,dt\right| = \left|\int_{E(f\neq g)} f(t)\varphi_n(t)\,dt\right| \leqslant K\varepsilon,$$

where the integral $\int_a^b g\varphi_n\,dt$ will be less than ε for sufficiently large n; for such n, we have

$$\left|\int_a^b f(t)\varphi_n(t)\,dt\right| < (K+1)\varepsilon,$$

which proves the theorem.

Example. Suppose $\varphi_n(t) = \cos nt$. Then

$$\int_a^c \varphi_n(t)\,dt = \frac{\sin nc - \sin na}{n} \to 0$$

and, obviously, both conditions of Lebesgue's theorem are satisfied. The case $\varphi_n(t) = \sin nt$ is considered analogously. We have thus proved

THEOREM 2 (RIEMANN-LEBESGUE). *For an arbitrary summable function defined on* $[a, b]$,

$$\lim_{n\to\infty} \int_a^b f(t)\cos nt\,dt = \lim_{n\to\infty} \int_a^b f(t)\sin nt\,dt = 0.$$

In particular, the Fourier coefficients

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt\,dt, \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt\,dt$$

of an arbitrary summable function tend to zero as[1] $n \to \infty$.

If the relation (7) holds for every summable function $f(t)$ defined on $[a, b]$, then we shall say the sequence $\{\varphi_n(t)\}$ *converges weakly to zero.*[2]

§ 2. REPRESENTATION OF A FUNCTION BY A SINGULAR INTEGRAL AT A GIVEN POINT

In the following discussion, if it is not explicitly stated so otherwise, we shall assume that the kernel $\Phi_n(t, x)$ is bounded for fixed n and x. Then the singular integral

$$f_n(x) = \int_a^b \Phi_n(t, x) f(t)\,dt$$

is meaningful for an arbitrary summable function $f(t)$.

THEOREM 1 (H. LEBESGUE). *If for fixed* $x\,(a < x < b)$ *and arbitrary* $\delta > 0$, *the kernel* $\Phi_n(t, x)$ *converges weakly to zero in each of the intervals* $[a, x-\delta]$, $[x+\delta, b]$ *and, furthermore,*

$$\int_a^b |\Phi_n(t, x)|\,dt < H(x),$$

where $H(x)$ *is independent of* n, *then the equality*

$$\lim_{n\to\infty} f_n(x) = f(x)$$

is valid for any summable function $f(t)$ *continuous at the point* x.

Proof. Since $\Phi_n(t, x)$ is a kernel,

---

[1] For *square-summable* functions, this proposition is an obvious consequence of the convergence the series $\sum(a_n^2 + b_n^2)$.

[2] We have deviated somewhat from the usual functional analysis terminology here.

$$\lim_{n\to\infty}\int_a^b \Phi_n(t,x)\,dt = 1,$$

and it will suffice to show that

$$\lim_{n\to\infty}\int_a^b [f(t)-f(x)]\,\Phi_n(t,x)\,dt = 0.$$

For this purpose, having taken $\varepsilon > 0$ we find a $\delta > 0$ such that

$$|f(t)-f(x)| < \frac{\varepsilon}{3H(x)}$$

for $|t-x| < \delta$.

Then for arbitrary n,

$$\left|\int_{x-\delta}^{x+\delta} [f(t)-f(x)]\,\Phi_n(t,x)\,dt\right| < \frac{\varepsilon}{3}.$$

But each of the integrals

$$\int_a^{x-\delta} [f(t)-f(x)]\,\Phi_n(t,x)\,dt, \quad \int_{x+\delta}^b [f(t)-f(x)]\,\Phi_n(t,x)\,dt$$

tends to zero as $n\to\infty$, so that for $n > n_0$, each of them will be less than $\frac{\varepsilon}{3}$ in absolute value; for these n, obviously

$$\left|\int_a^b [f(t)-f(x)]\,\Phi_n(t,x)\,dt\right| < \varepsilon,$$

which was to be proved.

This theorem is relevant to the representation of a summable function at points of continuity, but in general a summable function does not have a single point of continuity, which, of course, diminishes the interest of this theorem.

The question of representing a summable function at those points where this function is the derivative of its indefinite integral, or at Lebesgue points, is of great interest because, as we know, either of these sets of points fills almost the entire interval on which the function is defined.[3]

We now go over to this question.

LEMMA (I. P. NATANSON). *Let a summable function $f(t)$ possessing the property*

$$M = \sup_{0<h\leqslant b-a}\left\{\frac{1}{h}\left|\int_a^{a+h} f(t)\,dt\right|\right\} < +\infty \tag{1}$$

be defined on the closed interval $[a,b]$.

---

[3]Another type of "regular" point for a summable (and, in general, measurable) function is a point of approximate continuity. However, these points are not of great interest in the theory of singular integrals because, as is shown in I. P. Natanson: *Sur la représentation des fonctions aux points de continuité approximative par des intégrales singulières*, Fund. Math., Vol. XVIII, 1932, pp. 99-109, no singular integrals exist which represent an arbitrary summable function in all of its points of approximate continuity.

For any nonnegative decreasing function $g(t)$, defined and summable on $[a,b]$, the integral

$$\int_a^b f(t)g(t)\,dt \tag{2}$$

exists (it might be improper for $t = a$) and the inequality

$$\left|\int_a^b f(t)g(t)\,dt\right| \leqslant M\int_a^b g(t)\,dt \tag{3}$$

is valid.

To clarify the conditions of the lemma, we note that the case where $g(a) = +\infty$ is not excluded. If, however, $g(a) < +\infty$, the function $g(t)$ is bounded and integral (2) exists as an ordinary Lebesgue integral.

Going over to the proof of the lemma, we note that, without loss of generality, we may assume $g(b) = 0$. In fact, if this were not so, we would simply introduce the function $g^*(t)$ defined by the equalities

$$g^*(t) = \begin{cases} g(t), & \text{if } a \leqslant t < b \\ 0, & \text{if } t = b \end{cases}$$

to replace $g(t)$.

Having proved the theorem for $g^*(t)$, we could then replace $g^*(t)$ by $g(t)$ everywhere because this would have no effect on the value of the integrals involved. Thus, we shall assume $g(b) = 0$.

Let $a < \alpha < b$. On the closed interval $[\alpha, b]$, the function $g(t)$ is bounded and the integral

$$\int_a^b f(t)g(t)\,dt \tag{4}$$

certainly exists. If we set

$$F(t) = \int_a^t f(u)\,du,$$

integral (4) can be expressed in the form of a Stieltjes integral

$$\int_a^b f(t)g(t)\,dt = \int_a^b g(t)\,dF(t),$$

from which, after integrating by parts, we find

$$\int_a^b f(t)g(t)\,dt = -F(\alpha)g(\alpha) + \int_a^b F(t)\,d[-g(t)].$$

But, by (1), we have that

$$|F(t)| \leqslant M(t-a), \tag{5}$$

and since $g(t)$ decreases,

$$g(\alpha)(\alpha - a) \leqslant \int_a^\alpha g(t)\,dt. \tag{6}$$

This means that

$$|F(\alpha)g(\alpha)| \leqslant M\int_a^{\alpha} g(t)\,dt.$$

On the other hand, the function $-g(t)$ increases. This and (5) imply that

$$\left|\int_a^b F(t)\,d[-g(t)]\right| \leqslant M\int_a^b (t-a)\,d[-g(t)].$$

We transform the integral in the right member by using the formula for integration by parts:

$$\int_a^b (t-a)\,d[-g(t)] = g(\alpha)(\alpha-a) + \int_a^b g(t)\,dt.$$

From this, together with (6), it follows that

$$\left|\int_a^b (t-a)\,d[-g(t)]\right| \leqslant \int_a^b g(t)\,dt.$$

Combining the above, we obtain

$$\left|\int_a^b f(t)g(t)\,dt\right| \leqslant M\left\{\int_a^{\alpha} g(t)\,dt + \int_a^b g(t)\,dt\right\}. \tag{7}$$

Although this inequality was established under the assumption that $g(b)=0$, it also remains valid without this hypothesis as was already explained. This means we can replace the limit b by β, where $\alpha<\beta<b$. But then, letting α and β tend to a, we see that

$$\lim \int_a^{\beta} f(t)g(t)\,dt = 0,$$

which proves the existence of integral (2). Finally, we pass to the limit as $\alpha\to a$ in (7) and obtain (3); this completes the proof of the lemma.[4]

THEOREM 2 (P. I. ROMANOVSKY). *Suppose the kernel $\Phi_n(t,x)$ is positive and possesses the following property: for fixed n and x, the kernel $\Phi_n(t,x)$, as a function of t only, increases on the closed interval $[a,x]$ and decreases on the closed interval $[x,b]$.*

Then, for an arbitrary summable function $f(t)$ which is the derivative of its indefinite integral at the point x, we have

$$\lim_{n\to\infty} \int_a^b f(t)\Phi_n(t,x)\,dt = f(x). \tag{8}$$

Proof. Since $\Phi_n(t,x)$ is a kernel, it will suffice to verify that

$$\lim_{n\to\infty} \int_a^b [f(t)-f(x)]\,\Phi_n(t,x)\,dt = 0. \tag{9}$$

[4]In estimate (3), it is impossible to decrease the multiplier M since equality is attained in (3) for $f(t)=1$.

Splitting the last integral into two integrals extending over the closed intervals $[a, x]$ and $[x, b]$, we shall consider only the second of them since the first one is investigated analogously.

We take $\varepsilon > 0$ and find a $\delta > 0$ such that for $0 < h < \delta$,

$$\left| \frac{1}{h} \int_x^{x+h} [f(t) - f(x)]\, dt \right| < \varepsilon,$$

which is possible since $f(t)$ is the derivative of its indefinite integral at the point $t = x$.

Then according to the preceding lemma,

$$\left| \int_x^{x+\delta} [f(t) - f(x)]\, \Phi_n(t, x)\, dt \right| \leqslant \varepsilon \int_x^{x+\delta} \Phi_n(t, x)\, dt \leqslant \varepsilon \int_a^b \Phi_n(t, x)\, dt.$$

Since $\Phi_n(t, x)$ is a kernel,

$$\lim_{n \to \infty} \int_a^b \Phi_n(t, x)\, dt = 1.$$

This quantity is bounded since it has a finite limit. This means that there exists a constant $K(x)$ such that

$$\int_a^b \Phi_n(t, x)\, dt < K(x).$$

Hence,

$$\left| \int_x^{x+\delta} [f(t) - f(x)]\, \Phi_n(t, x)\, dt \right| < \varepsilon K(x). \tag{10}$$

On the other hand, if $x + \delta \leqslant t \leqslant b$,

$$\Phi_n(t, x) \leqslant \Phi_n(x + \delta, x) \leqslant \frac{1}{\delta} \int_x^{x+\delta} \Phi_n(t, x)\, dt < \frac{K(x)}{\delta}.$$

This means the functions $\varphi_n(t) = \Phi_n(t, x)$ are uniformly bounded on the closed interval $[x + \delta, b]$, and condition (5) of Lebesgue's theorem, §1, is satisfied. But the second condition of this theorem, i.e. (6), is also satisfied for these functions because $\Phi_n(t, x)$ is a kernel. Hence $\Phi_n(t, x)$ converges weakly to zero on the closed interval $[x + \delta, b]$ and for sufficiently large n,

$$\left| \int_{x+\delta}^b [f(t) - f(x)]\, \Phi_n(t, x)\, dt \right| < \varepsilon.$$

For these n,

$$\left| \int_x^b [f(t) - f(x)]\, \Phi_n(t, x)\, dt \right| < \varepsilon [K(x) + 1]$$

[see (10)], so that

$$\lim_{n\to\infty}\int_x^b [f(t)-f(x)]\,\Phi_n(t,x)\,dt=0. \tag{11}$$

This completes the proof of the theorem.

As an example of its application, we take the *Weierstrass integral*

$$W_n(x)=\frac{n}{\sqrt{\pi}}\int_a^b e^{-n^2(t-x)^2}f(t)\,dt.$$

The function

$$W_n(t,x)=\frac{n}{\sqrt{\pi}}\,e^{-n^2(t-x)^2}$$

is a kernel because, for $\alpha<x<\beta$,

$$\int_a^\beta W_n(t,x)\,dt=\frac{1}{\sqrt{\pi}}\int_{n(a-x)}^{n(\beta-x)} e^{-z^2}dz\to\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty} e^{-z^2}dz=1$$

This function is positive; it increases for $a\leqslant t\leqslant x$ and decreases for $x\leqslant t\leqslant b$. This means that for every $f(t)\in L$,

$$\lim_{n\to\infty} W_n(x)=f(x)$$

at every point x where $f(t)$ is the derivative of its indefinite integral.

We agree to say that the function $\Psi(t,x)$ is a *convex majorant* of the function $\Phi(t,x)$ if

$$|\Phi(t,x)|\leqslant\Psi(t,x)$$

and if, for fixed x, $\Psi(t,x)$ increases on the closed interval $[a,x]$ and decreases on the closed interval $[x,b]$.

THEOREM 3 (D. K. FADDEYEV). *If for every n, the kernel $\Phi_n(t,x)$ has a convex majorant $\Psi_n(t,x)$ such that*

$$\int_a^b \Psi_n(t,x)\,dt<K(x)<+\infty,$$

where $K(x)$ depends only on x, then (8) *is valid for arbitrary $f(t)\in L$, having the point $t=x$ as a Lebesgue point.*

Proof. Here, also, it will suffice to prove equality (11). Taking an $\varepsilon>0$, we find a $\delta>0$ such that for $0<h\leqslant\delta$,

$$\frac{1}{h}\int_x^{x+h}|f(t)-f(x)|\,dt<\varepsilon.$$

According to the lemma,

$$\left|\int_x^{x+\delta}\{f(t)-f(x)\}\,\Phi_n(t,x)\,dt\right|\leqslant\int_x^{x+h}|f(t)-f(x)|\cdot\Psi_n(t,x)\,dt\leqslant\varepsilon\int_x^{x+h}\Psi_n(t,x)\,dt<\varepsilon K(x).$$

On the other hand, on the closed interval $[x+\delta,b]$, the sequence $\{\varphi_n(t)\}=\{\Phi_n(t,x)\}$ converges weakly to zero because, for $t\in[x+\delta,b]$,

$$|\Phi_n(t,x)| \leqslant \Psi_n(t,x) \leqslant \Psi_n(x+\delta,x) \leqslant \frac{1}{\delta}\int_x^{x+\delta} \Psi_n(t,x)\,dt < \frac{K(x)}{\delta}.$$

Noting this, we complete the proof of Faddeyev's theorem exactly as above.[5]

§ 3. APPLICATION TO THE THEORY OF FOURIER SERIES

We have already defined the concept of the Fourier series of a function $f(x)$ in an arbitrary orthonormal system $\{\omega_k(x)\}$ (Chapter VII, §3). In particular, if we are dealing with the trigonometric system

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \ldots, \tag{1}$$

then the Fourier series of the function $f(x)$ will be the series

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx), \tag{2}$$

where

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos kx\,dx, \quad b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin kx\,dx. \tag{3}$$

In Chapter VII, we assumed that $f(x) \in L_2$. This assumption assured the existence of the Fourier coefficients

$$c_k = \int_a^b f(x)\,\omega_k(x)\,dx$$

of the function $f(x)$ in an arbitrary orthonormal system. But the functions of system (1) are bounded. Hence, the coefficients (3), and consequently series (2), can be formed for an arbitrary summable function.

The question of convergence of series (2) reduces to the investigation of a singular integral. In fact, if

$$S_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx),$$

then, by (3),

$$S_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} \left[\frac{1}{2} + \sum_{k=1}^{n} \cos k(t-x)\right] f(t)\,dt.$$

---

[5] D. K. Faddeyev: *On the representation of summable functions by means of singular integrals at Lebesgue points*, Mat. Sbornik, Vol. 1 (43), No. 3, 1936, pp. 351-368, proved that the conditions of the theorem are *necessary* in order that (8) hold for every $f(t) \in L$ having a Lebesgue point at $t = x$.

Using the well-known[6] formula

$$\frac{1}{2}+\sum_{k=1}^{n}\cos k\alpha=\frac{\sin\dfrac{2n+1}{2}\alpha}{2\sin\dfrac{\alpha}{2}}, \tag{4}$$

the sum $S_n(x)$ can be written in the form

$$S_n(x)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sin\dfrac{2n+1}{2}(t-x)}{\sin\dfrac{t-x}{2}}f(t)\,dt. \tag{5}$$

This integral is the *Dirichlet singular integral.*

We shall not stop to consider questions of convergence of the series (2); the interested reader is referred to the exhaustive treatise by Zygmund.[7] On the other hand, we shall consider the question of *summing series* (2) *by the Cesàro method.* This method consists in finding the limit of the arithmetic mean of the first n partial sums $S_k(x)$:

$$\sigma_n(x)=\frac{S_0(x)+S_1(x)+\ldots+S_{n-1}(x)}{n}. \tag{6}$$

It is known that in the case where the series (2) converges at the point x, the sequence $\{\sigma_n(x)\}$ converges to the sum of the series, but this sequence may also converge when series (2) diverges.[8]

To investigate $\sigma_n(x)$, we transform it with the aid of formula (5):

$$\sigma_n(x)=\frac{1}{2n\pi}\int_{-\pi}^{\pi}\left[\sum_{k=0}^{n-1}\sin\frac{2k+1}{2}(t-x)\right]\frac{f(t)}{\sin\dfrac{t-x}{2}}\,dt.$$

But

$$\sum_{k=0}^{n-1}\sin(2k+1)\alpha=\frac{\sin^2 n\alpha}{\sin\alpha}. \tag{7}$$

---

[6]This formula is easily derived. To do this, we add the equalities

$$\sin(k+\tfrac{1}{2})\alpha-\sin(k-\tfrac{1}{2})\alpha=2\sin\frac{\alpha}{2}\cos k\alpha\,(k=1,2,\ldots,n)$$

$$\sin\frac{\alpha}{2}=\sin\frac{\alpha}{2}.$$

We obtain

$$\sin(n+\tfrac{1}{2})\alpha=2\sin\frac{\alpha}{2}\left[\frac{1}{2}+\sum_{k=1}^{n}\cos k\alpha\right],$$

from which (4) follows.

[7]A. Zygmund: *Trigonometrical Series*, New York, Chelsea, 1952.

[8]See, for example, I. I. Privalov: *Fourier Series*, ONTI, 1934; L. V. Kantorovič: *Definite Integrals and Fourier Series*, LGU Press, 1940; G. M. Fichtenholz: *Course in Differential and Integral Calculus*, Vol. 3, GTTI, 1949; or A. Zygmund: *Trigonometrical Series*, Chapter III.

In fact, adding the equalities

$$\cos 2k\alpha - \cos 2(k+1)\alpha = 2\sin\alpha \sin(2k+1)\alpha \quad (k = 0, 1, \ldots, n-1),$$

we find

$$2\sin\alpha \sum_{k=0}^{n-1} \sin(2k+1)\alpha = 1 - \cos 2n\alpha = 2\sin^2 n\alpha,$$

from which (7) follows.

With the aid of (7), we obtain

$$\sigma_n(x) = \frac{1}{2n\pi} \int_{-\pi}^{\pi} \left[\frac{\sin n\dfrac{t-x}{2}}{\sin\dfrac{t-x}{2}} \right]^2 f(t)\,dt. \tag{8}$$

Integral (8) is the *Fejér singular integral.* We shall show that the conditions of the Faddeyev theorem are satisfied for it.

To this end, consider the function $f(t) = 1$. Calculating its Fourier coefficients by formulas (3), we obtain

$$a_0 = 2, \quad a_k = b_k = 0 \qquad (k = 1, 2, \ldots).$$

This means that for this function,

$$S_n(x) = 1 \qquad (n = 0, 1, 2, \ldots),$$

and consequently, that also

$$\sigma_n(x) = 1.$$

But expressing $\sigma_n(x)$ by the Fejér integral, we obtain that

$$\frac{1}{2n\pi} \int_{-\pi}^{\pi} \left[\frac{\sin n\dfrac{t-x}{2}}{\sin\dfrac{t-x}{2}} \right]^2 dt = 1. \tag{9}$$

Having noted this, consider the point $x \in (-\pi, \pi)$. Let $-\pi \leqslant \alpha < x < \beta \leqslant \pi$. If $t \in [-\pi, \alpha]$, then

$$\frac{1}{\sin^2\dfrac{t-x}{2}} \leqslant \max\left\{ \frac{1}{\sin^2\dfrac{\alpha-x}{2}}, \frac{1}{\sin^2\dfrac{-\pi-x}{2}} \right\} = A(x, \alpha),$$

and, consequently,

$$\frac{1}{2n\pi} \int_{-\pi}^{\alpha} \left[\frac{\sin n\dfrac{t-x}{2}}{\sin\dfrac{t-x}{2}} \right]^2 dt < \frac{A(x,\alpha)}{n},$$

where $A(x, \alpha)$ does not depend on n. It follows that

$$\lim_{n\to\infty} \frac{1}{2n\pi} \int_{-\pi}^{\alpha} \left[\frac{\sin n\dfrac{t-x}{2}}{\sin\dfrac{t-x}{2}} \right]^2 dt = 0.$$

We can verify analogously that the integral over the closed interval $[\beta, \pi]$ tends to zero. Combining this with (9), we find that

$$\lim_{n\to\infty} \frac{1}{2n\pi} \int_a^\beta \left[\frac{\sin n\dfrac{t-x}{2}}{\sin \dfrac{t-x}{2}} \right]^2 dt = 1,$$

so that the function

$$\frac{1}{2n\pi} \left[\frac{\sin n\dfrac{t-x}{2}}{\sin \dfrac{t-x}{2}} \right]^2$$

is a kernel.

It is easy to construct a convex majorant for this kernel. To do this, note that $|\sin z| \leqslant |z|$. Hence

$$\frac{1}{\sin^2 z} \geqslant \frac{1}{z^2}.$$

But

$$\frac{1}{\sin^2 z} \geqslant 1.$$

Therefore

$$\frac{1}{\sin^2 z} \geqslant \frac{1}{2}\left(1 + \frac{1}{z^2}\right) = \frac{z^2+1}{2z^2}$$

and

$$\sin^2 \frac{n(t-x)}{2} \leqslant \frac{2n^2(t-x)^2}{n^2(t-x)^2+4}. \tag{10}$$

On the other hand, when $|z| \leqslant \frac{\pi}{2}$, then $|\sin z| \geqslant \frac{2}{\pi}|z|$, so that[9]

$$\sin^2 \frac{t-x}{2} \geqslant \frac{1}{\pi^2}(t-x)^2. \tag{11}$$

---

[9]Since $-\pi < x < \pi$, $-\pi \leqslant t \leqslant \pi$, $\left|\dfrac{t-x}{2}\right|$ may also be larger than $\frac{\pi}{2}$. However, this is not essential. In fact, set $a = \max\{-\pi, x-\pi\}$, $b = \min\{\pi, x+\pi\}$. It is not difficult to see that the difference between the Fejér integral (8) and the integral

$$\sigma_n^* = \frac{1}{2n\pi} \int_a^b \left[\frac{\sin n\dfrac{t-x}{2}}{\sin \dfrac{t-x}{2}} \right]^2 f(t)\,dt$$

tends to zero as n increases (because, for example, $\left|\sin \dfrac{t-x}{2}\right| \geqslant \left|\sin \dfrac{-\pi-x}{2}\right|$ for $-\pi \leqslant t \leqslant a$); therefore, all the above discussion is applicable to the integral σ_n^*.

It follows from (10) and (11) that

$$\frac{1}{2n\pi}\left[\frac{\sin n\frac{t-x}{2}}{\sin\frac{t-x}{2}}\right]^2 \leqslant \frac{1}{2n\pi}\,\frac{2n^2(t-x)^2}{n^2(t-x)^2+4}\cdot\frac{\pi^2}{(t-x)^2} = \frac{n\pi}{n^2(t-x)^2+4}\,.$$

The function $\dfrac{n\pi}{n^2(t-x)^2+4}$ is a convex majorant for the Fejér kernel. But

$$\int_{-\pi}^{\pi}\frac{n\pi\,dt}{n^2(t-x)^2+4} < \int_{-\infty}^{\infty}\frac{\pi dz}{z^2+4} = \frac{\pi^2}{2}\,,$$

i.e. the integrals of the majorant are bounded by a number which does not depend on n.

Hence, the Fejér integral satisfies the conditions of Faddeyev's theorem. This implies the next theorem.

THEOREM 1 (L. FEJÉR-H. LEBESGUE). *Almost everywhere on* $[-\pi, +\pi]$,

$$\lim_{n\to\infty}\sigma_n(x) = f(x). \tag{12}$$

This relation is satisfied at all Lebesgue points and hence, a fortiori, at all points of continuity of the function $f(t)$, *lying interior to* $[-\pi, +\pi]$.

In Chapter VII, we established that the trigonometric system (1) is *complete*. This meant that every function $f(x)\in L_2$ for which all Fourier coefficients (3) vanish, is equivalent to zero. Now we can liberate ourselves from the restriction that $f(x)$ be square-summable. The following theorem is valid.

THEOREM 2. *If all the Fourier coefficients* (3) *of a summable function* $f(x)$ *equal zero, then* $f(x)$ *is equivalent to zero.*

In fact, in this case $\sigma_n(x) = 0$ and, consequently, $f(x) = 0$ at all points where (12) holds, i.e. almost everywhere.

Theorem 1 enables us to make certain assertions concerning the behavior of the partial sums $S_n(x)$.

In this connection, note that

$$\sigma_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n-1}\frac{n-k}{n}\,(a_k\cos kx + b_k\sin kx),$$

so that

$$S_n(x) - \sigma_n(x) = \sum_{k=1}^{n}\frac{k}{n}\,(a_k\cos kx + b_k\sin kx). \tag{13}$$

It follows that

$$\frac{1}{\pi}\int_{-\pi}^{\pi}[S_n(x)-\sigma_n(x)]^2\,dx = \sum_{k=1}^{n}\frac{k^2}{n^2}\,(a_k^2+b_k^2). \tag{14}$$

Consider the sequence $n_1, n_2, n_3, \ldots$ of natural numbers, for which

$$\frac{n_{i+1}}{n_i} > A+1. \tag{15}$$

Such a sequence is called a *lacunary sequence.* The following theorem is valid.

THEOREM 3 (A. N. KOLMOGOROV). *Let $f(x)$ be square-summable. If $\{n_i\}$ is a lacunary sequence, then*

$$\lim_{i\to\infty} S_{n_i}(x) = f(x)$$

almost everywhere on $[-\pi, +\pi]$.

Proof. By the Fejér-Lebesgue theorem it will suffice to show that

$$\lim_{i\to\infty} [S_{n_i}(x) - \sigma_{n_i}(x)] = 0$$

almost everywhere.

For this (Theorem 11, §1, Chapt. VI, Corollary), it suffices to show that

$$\sum_{i=1}^{\infty} \int_{-\pi}^{\pi} (S_{n_i} - \sigma_{n_i})^2 \, dx < +\infty,$$

or, equivalently, that

$$Q = \sum_{i=1}^{\infty} \left[\frac{1}{n_i^2} \sum_{k=1}^{n_i} k^2 (a_k^2 + b_k^2) \right] < +\infty.$$

The sum Q can be written as

$$\begin{aligned} Q = {} & \frac{1}{n_1^2} \sum_{k=1}^{n_1} u_k + \\ & + \frac{1}{n_2^2} \sum_{k=1}^{n_1} u_k + \frac{1}{n_2^2} \sum_{k=n_1+1}^{n_2} u_k + \\ & + \frac{1}{n_3^2} \sum_{k=1}^{n_1} u_k + \frac{1}{n_3^2} \sum_{k=n_1+1}^{n_2} u_k + \frac{1}{n_3^2} \sum_{k=n_2+1}^{n_3} u_k + \\ & + \ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

where $u_k = k^2 (a_k^2 + b_k^2)$. Summing these series by columns, we obtain

$$Q = \sum_{i=1}^{\infty} \left(\sum_{s=i}^{\infty} \frac{1}{n_s^2} \right) \left(\sum_{k=n_{i-1}+1}^{n_i} u_k \right) \qquad (n_0 = 0).$$

But since

$$\frac{n_i}{n_s} \leqslant \frac{1}{A^{s-i}},$$

we have

$$\left(\sum_{s=i}^{\infty} \frac{1}{n_s^2} \right) \left(\sum_{k=n_{i-1}+1}^{n_i} u_k \right) < \sum_{s=i}^{\infty} \left(\frac{n_i}{n_s} \right)^2 \cdot \sum_{n_{i-1}+1}^{n_i} (a_k^2 + b_k^2) <$$

$$< \sum_{s=0}^{\infty} \frac{1}{A^{2s}} \cdot \sum_{k=n_{i-1}+1}^{n_i} (a_k^2 + b_k^2) = \frac{A^2}{A^2 - 1} \sum_{k=n_{i-1}+1}^{n_i} (a_k^2 + b_k^2).$$

Hence,

$$Q < \frac{A^2}{A^2-1}\sum_{i=1}^{\infty}\sum_{k=n_{i-1}+1}^{n_i}(a_k^2+b_k^2) = \frac{A^2}{A^2-1}\sum_{k=1}^{\infty}(a_k^2+b_k^2) < +\infty,$$

because the series $\sum (a_k^2+b_k^2)$ converges. This completes the proof of the theorem.

It is useful to note how closely related this theorem of A. N. Kolmogorov is to the Kaczmarz and Rademacher theorems of §3, Chapt. VII. Namely, in both cases indices n_i are sought so that the partial sums $S_{n_i}(x)$ of the Fourier series $\sum_{k=1}^{\infty} c_k\omega_k(x)$ of the function $f(x)$ converge to $f(x)$ almost everywhere. However, in the Kaczmarz and Rademacher theorems, it is assumed the coefficients c_k are given and the indices n_i are determined by them independently of the system $\{\omega_k(x)\}$. In contrast, A. N. Kolmogorov considers a *definite* trigonometric orthonormal system and characterizes the indices n_i independently of the coefficients c_k.

DEFINITION. The trigonometric series

$$\sum_{i=1}^{\infty}(a_{n_i}\cos n_i x + b_{n_i}\sin n_i x)$$

is said to be *lacunary* if $\{n_i\}$ is a lacunary sequence.

THEOREM 4 (A. N. KOLMOGOROV). *If the Fourier series of a summable function $f(x)$ is lacunary, then it converges to $f(x)$ almost everywhere.*

Proof. If $n_i \leqslant n < n_{i+1}$, then

$$S_n(x) = S_{n_i}(x) \text{ because } a_{n_i+1} = b_{n_i+1} = a_{n_i+2} = \ldots = b_n = 0.$$

Therefore, it will suffice to show that

$$\lim_{i\to\infty} S_{n_i}(x) = f(x) \qquad (*)$$

almost everywhere, and this assertion, in its turn, will be proved if we establish that

$$\lim_{i\to\infty}[S_{n_i}(x) - \sigma_{n_i}(x)] = 0 \qquad (*_*)$$

almost everywhere.

We note, by the way, that for the particular case when $f(x) \in L_2$, the equality (*), and with it also the theorem, follows at once from the preceding theorem of Kolmogorov.

Going over to the proof of relation (*$_*$), we note that by (13)

$$|S_{n_i}(x) - \sigma_{n_i}(x)| \leqslant \sum_{k=1}^{n_i}\frac{k}{n_i}(|a_k| + |b_k|).$$

If k does not coincide with any of the numbers $n_1, n_2, n_3, \ldots$, then

$$a_k = b_k = 0.$$

Hence,

$$|S_{n_i}(x) - \sigma_{n_i}(x)| \leqslant \sum_{m=1}^{i}\frac{n_m}{n_i}(|a_{n_m}| + |b_{n_m}|).$$

By virtue of the Riemann-Lebesgue theorem, §1,

$$\lim_{m\to\infty} a_{n_m} = \lim_{m\to\infty} b_{n_m} = 0.$$

Noting this, we take $\varepsilon > 0$ and find an m_0 such that for $m > m_0$,

$$|a_{n_m}| + |b_{n_m}| < \varepsilon.$$

Then, denoting

$$\sum_{m=1}^{m_0} n_m (|a_{n_m}| + |b_{n_m}|)$$

by M, we will have, for $i > m_0$, that

$$|S_{n_i}(x) - \sigma_{n_i}(x)| < \frac{M}{n_i} + \varepsilon \sum_{m=m_0+1}^{i} \frac{n_m}{n_i}. \tag{16}$$

But since

$$\frac{n_m}{n_i} \leqslant \frac{1}{A^{i-m}}, \text{ we have}$$

$$\sum_{m=m_0+1}^{i} \frac{n_m}{n_i} < \sum_{m=m_0+1}^{i} \frac{1}{A^{i-m}} < \sum_{k=0}^{\infty} \frac{1}{A^k} = \frac{A}{A-1}.$$

From this and (16) it follows that

$$|S_{n_i}(x) - \sigma_{n_i}(x)| < \frac{M}{n_i} + \frac{A}{A-1}\varepsilon.$$

If i is sufficiently large, $\frac{M}{n_i} < \varepsilon$, and for these i,

$$|S_{n_i}(x) - \sigma_{n_i}(x)| < \left(1 + \frac{A}{A-1}\right)\varepsilon,$$

which proves the theorem.

Remark. The Abel-Poisson method of summing a Fourier series is also well known.[10] It consists in forming the auxiliary series

$$S_r(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} r^k (a_k \cos kx + b_k \sin kx),$$

which converges for $0 < r < 1$. If there exists a finite limit

$$\lim_{r \to 1-0} S_r(x),$$

this limit is taken as the generalized sum of the series

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx). \tag{17}$$

It is not difficult to show that at every point x at which the series (17) is summable by the Cesàro-Fejér method, it is also summable by the Abel-Poisson method[11] to the

---

[10] I. I. Privalov: *Fourier Series*, p. 108; L. V. Kantorovič: *Definite Integrals and Fourier Series*, p. 217; G. M. Fichtenholz: *Course in Differential and Integral Calculus*, Vol. 3, p. 723; A. Zygmund: *Trigonometrical Series*, p. 50.

[11] I. I. Privalov: *Fourier Series*, p. 110; L. V. Kantorovič: *Definite Integrals and Fourier Series*, p. 218; G. M. Fichtenholz: *Course in Differential and Integral Calculus*, Vol. 3, p. 731; A. Zygmund: *Trigonometrical Series*, Chapter III.

same sum. This means *the Fourier series of a summable function $f(x)$ can be summed to the function $f(x)$ by the Abel-Poisson method almost everywhere.* This fact can also be proved not using the Fejér-Lebesgue theorem, but rather using P. I. Romanovsky's [12] theorem in the investigation of the singular integral in terms of which $S_r(x)$ is expressed:

$$S_r(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \frac{1 - r^2}{1 - 2r\cos(t - x) + r^2} dt$$

(Poisson integral). We shall not take time to consider this.

§ 4. FURTHER PROPERTIES OF TRIGONOMETRIC AND FOURIER SERIES

Not every trigonometric series

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx)$$

is the Fourier series for a summable function, even if its coefficients tend to zero. In order to verify this, we shall establish some assertions which are also of interest in themselves.

LEMMA 1 (N. ABEL). *Let $a_1, a_2, \ldots, a_n$ be given numbers. Set*

$$s_k = \sum_{i=1}^{k} a_i$$

and let

$$|s_k| \leqslant A \qquad (k = 1, 2, \ldots, n).$$

If

$$q_1 > q_2 > \ldots > q_n > 0,$$

then

$$\left| \sum_{k=1}^{n} a_k q_k \right| \leqslant A q_1.$$

Proof. If $k > 1$, then $a_k = s_k - s_{k-1}$. This means that

$$\sum_{k=1}^{n} a_k q_k = s_1 q_1 + \sum_{k=2}^{n} (s_k - s_{k-1}) q_k = \sum_{k=1}^{n} s_k q_k - \sum_{k=2}^{n} s_{k-1} q_k$$

and hence,

$$\sum_{k=1}^{n} a_k q_k = \sum_{k=1}^{n-1} s_k (q_k - q_{k+1}) + s_n q_n.$$

Thus,

$$\left| \sum_{k=1}^{n} a_k q_k \right| \leqslant A \left[\sum_{k=1}^{n-1} (q_k - q_{k+1}) + q_n \right] = A q_1,$$

which was to be proved.

---

[12] The method based on Romanovsky's theorem gives a stronger result than the Fejér-Lebesgue theorem. Namely, it shows that the Fourier series of a function $f(x)$ is Abel-Poisson summable at every point where $f(x)$ is the derivative of its indefinite integral (and not only at Lebesgue points, as follows from the Fejér-Lebesgue theorem).

DEFINITION. We say the series

$$a_1 + a_2 + a_3 + \dots$$

satisfies the *Abel condition* if

$$|\sum_{k=1}^{n} a_k| \leqslant A \qquad (n = 1, 2, 3, \dots).$$

LEMMA 2. *For fixed x, each of the series*

$$\cos x + \cos 2x + \cos 3x + \dots \qquad (x \neq 2k\pi)$$
$$\sin x + \sin 2x + \sin 3x + \dots \qquad (x \text{ arbitrary}),$$

satisfies the Abel condition.

Proof. Set

$$A_n = \sum_{k=1}^{n} \cos kx, \qquad B_n = \sum_{k=1}^{n} \sin kx.$$

To avoid repetition, we estimate these sums by a method different from that applied in analogous cases above.

These sums are the real and imaginary parts of the sum

$$C_n = \sum_{k=1}^{n} e^{kxi} = \frac{e^{xi} - e^{(n+1)xi}}{1 - e^{xi}} .$$

Hence, it will suffice to estimate the absolute value of the last sum.

Obviously,

$$|C_n| \leqslant \frac{2}{|1 - e^{xi}|} = \frac{1}{\left|\sin \frac{x}{2}\right|}.$$

Therefore

$$|A_n| \leqslant \frac{1}{\left|\sin \frac{x}{2}\right|}, \quad |B_n| < \frac{1}{\left|\sin \frac{x}{2}\right|}$$

and the lemma is proved for $x \neq 2k\pi$. If $x = 2k\pi$, the lemma is trivial for the sine series.

THEOREM 1 (N. ABEL). *Suppose the series $\sum_{k=1}^{\infty} a_k$ satisfies the Abel condition. If*

$$q_1 > q_2 > \dots > q_n > \dots; \quad \lim_{n \to \infty} q_n = 0,$$

then the series $\sum_{k=1}^{\infty} a_k q_k$ converges.

Proof. Set

$$S_n = \sum_{k=1}^{n} a_k q_k.$$

Then, by Abel's lemma, for $m > n$,

$$|S_m - S_n| = |\sum_{k=n+1}^{m} a_k q_k| \leqslant 2Aq_{n+1}$$

and for sufficiently large n, this quantity is arbitrarily small, from which the theorem follows.

COROLLARY. *If*

$$q_1 > q_2 > \dots > q_n > \dots, \quad \lim_{n \to \infty} q_n = 0,$$

then each of the series

$$\sum_{n=1}^{\infty} q_n \cos nx \qquad (x \neq 2k\pi), \qquad \sum_{n=1}^{\infty} q_n \sin nx \qquad (x \text{ arbitrary})$$

converges.

For example, *the series*

$$\sum_{n=2}^{\infty} \frac{\sin nx}{\ln n} \tag{1}$$

converges for every x. Further on, we shall show that (1) is not the Fourier series of any summable function. For this, the following lemma is needed.

LEMMA 3. *Let*

$$\psi_n(x) = \sum_{k=1}^{n} \frac{\sin kx}{k}.$$

The estimate

$$|\psi_n(x)| < 2\sqrt{\pi}$$

is valid for any x *and* n.

Proof. First of all, let $0 < x < \pi$. Let q be an integer such that

$$q \leqslant \frac{\sqrt{\pi}}{x} < q + 1.$$

Then

$$\psi_n(x) \leqslant |\psi_q(x)| + \left| \sum_{k=q+1}^{n} \frac{\sin kx}{k} \right|.$$

(If $q = 0$, the first term, and if $q \geqslant n$, the second term, of the right member vanishes.) Since $|\sin\alpha| \leqslant |\alpha|$,

$$|\psi_q(x)| \leqslant \sum_{k=1}^{q} \frac{|\sin kx|}{k} \leqslant qx \leqslant \sqrt{\pi}. \tag{2}$$

On the other hand, by Abel's lemma,

$$\left| \sum_{k=q+1}^{n} \frac{\sin kx}{k} \right| \leqslant \frac{A}{q+1},$$

where $A = \max \left| \sum_{k=q+1}^{i} \sin kx \right| \ (q+1 \leqslant i \leqslant n)$. But by a verbatim repetition of the line of reasoning followed in the proof of Lemma 2, we see that

$$\left| \sum_{k=q+1}^{i} \sin kx \right| \leqslant \frac{1}{\left| \sin \frac{x}{2} \right|};$$

this means that $A \leqslant \dfrac{1}{\sin \dfrac{x}{2}}$ and, consequently,

$$\left|\sum_{k=q+1}^{n}\frac{\sin kx}{k}\right| \leqslant \frac{1}{(q+1)\sin\frac{x}{2}}.$$

Taking into consideration that $\sin\frac{x}{2} \geqslant \frac{x}{\pi}$ and $q+1 > \frac{\sqrt{\pi}}{x}$, we have

$$\left|\sum_{k=q+1}^{n}\frac{\sin kx}{k}\right| \leqslant \frac{1}{\frac{\sqrt{\pi}}{x}\cdot\frac{x}{\pi}} = \sqrt{\pi},$$

from which, in connection with (2), the required estimate follows. Inasmuch as $|\psi_n(x)|$ is an even function, this estimate is also valid for $-\pi < x < 0$. For $x = 0$ and $x = \pi$ this estimate is trivial, i.e. it is true for $-\pi < x < \pi$ and then by virtue of the periodicity of $\psi_n(x)$, it is always true, which was to be proved.

THEOREM 2. *Let $f(x)$ be a summable function. If*

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx, \text{ then the series } \sum_{n=1}^{\infty}\frac{b_n}{n}$$

converges.[13]

Proof. The series $\sum_{n=1}^{\infty}\frac{\sin nx}{n}$ converges by the corollary to Theorem 1. If its sum is $\psi(x)$, then for all x,

$$\lim_{n\to\infty}\psi_n(x)f(x) = \psi(x)f(x).$$

By Lemma 3,

$$|\psi_n(x)f(x)| \leqslant 2\sqrt{\pi}|f(x)|.$$

This means, by Lebesgue's theorem on passing to the limit under the integral sign, that

$$\lim_{n\to\infty}\int_{-\pi}^{\pi}\psi_n(x)f(x)\,dx = \int_{-\pi}^{\pi}\psi(x)f(x)\,dx.$$

But

$$\frac{1}{\pi}\int_{-\pi}^{\pi}\psi_n(x)f(x)\,dx = \sum_{k=1}^{n}\frac{b_k}{k},$$

from which the theorem follows.

Turning to series (1), we see that the series

$$\sum_{n=2}^{\infty}\frac{1}{n\ln n}$$

[13] If $f(x)\in L_2$, this theorem is an obvious consequence of the convergence of both series $\sum b_n^2$ and $\sum\frac{1}{n^2}$ and of the inequality $\frac{|b_n|}{n} < b_n^2 + \frac{1}{n^2}$. In this case, the series $\sum\frac{a_n}{n}$ also converges, but it may diverge in the case $f(x)$ is an arbitrary summable function.

diverges.[14] This means (1) *is an example of a trigonometric series which converges everywhere but which is not the Fourier series of any summable function.*

Along the same order of ideas, we prove the following theorem.

THEOREM 3. *Let a summable function $f(x)$, defined on the interval $[-\pi, \pi]$, have the series*

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

as its Fourier series. If $[A, B] \subset [-\pi, \pi]$, then

$$\int_A^B f(x)\,dx = \int_A^B \frac{a_0}{2}\,dx + \sum_{n=1}^{\infty} \int_A^B (a_n \cos nx + b_n \sin nx)\,dx.$$

In other words, the Fourier series of a summable function can be *integrated term by term.* This fact is quite remarkable inasmuch as this series may not converge.

For the case when the function $f(x)$ is *square*-summable, the theorem follows immediately from the closure of the trigonometric system and the corollary to Theorem 1, §3, Chapter VII.

To prove the theorem in the general case, we set

$$\varphi(x) = \begin{cases} 1 \text{ for } x \in [A, B] \\ 0 \text{ for } x \overline{\in} [A, B]. \end{cases}$$

This function, as is known from elementary analysis, is expandable in the Fourier series

$$\varphi(x) = \frac{\alpha_0}{2} + \sum_{k=1}^{\infty} (\alpha_k \cos kx + \beta_k \sin kx)$$

at all points of the closed interval $[-\pi, \pi]$, except perhaps at the points $-\pi, A, B, \pi$. Let

$$S_n(x) = \frac{\alpha_0}{2} + \sum_{k=1}^{n} (\alpha_k \cos kx + \beta_k \sin kx).$$

This sum can be written in a different form if the coefficients α_k and β_k are actually calculated:

$$\alpha_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} \varphi(x)\,dx = \frac{B - A}{\pi},$$

---

[14] Since $\frac{1}{x \ln x}$ is a decreasing function,

$$\frac{1}{n \ln n} > \int_n^{n+1} \frac{dx}{x \ln x}.$$

Therefore,

$$\sum_{n=2}^{N} \frac{1}{n \ln n} > \int_2^N \frac{dx}{x \ln x} = \ln \ln N - \ln \ln 2.$$

$$\alpha_k = \frac{1}{\pi}\int_{-\pi}^{\pi} \varphi(x)\cos kx\,dx = \frac{\sin kB - \sin kA}{k\pi},$$

$$\beta_k = \frac{1}{\pi}\int_{-\pi}^{\pi} \varphi(x)\sin k\,x\,dx = \frac{\cos kA - \cos kB}{k\pi}.$$

Substituting these quantities into the expression for $S_n(x)$, we find

$$S_n(x) = \frac{B-A}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{n}\left[\frac{\sin k(B-x)}{k} - \frac{\sin k(A-x)}{k}\right].$$

Hence, by Lemma 3,

$$|S_n(x)| \leqslant \frac{B-A}{2\pi} + \frac{4}{\sqrt{\pi}}$$

for all x and n, i.e. the sums $S_n(x)$ are uniformly bounded. Using the Lebesgue theorem on passage to the limit under the integral sign, we find

$$\int_{-\pi}^{\pi} f(x)\varphi(x)\,dx = \lim_{n\to\infty}\int_{-\pi}^{\pi} f(x)S_n(x)\,dx,$$

which can also be written as

$$\int_A^B f(x)\,dx = \frac{\alpha_0}{2}\int_{-\pi}^{\pi} f(x)\,dx + \sum_{k=1}^{\infty}\left[\alpha_k\int_{-\pi}^{\pi} f(x)\cos kx\,dx + \beta_k\int_{-\pi}^{\pi} f(x)\sin kx\,dx\right]$$

or

$$\int_A^B f(x)\,dx = \frac{a_0}{2}(B-A) + \sum_{k=1}^{\infty}\left[a_k\frac{\sin kB - \sin kA}{k} + b_k\frac{\cos k\,A - \cos kB}{k}\right],$$

and this equality is equivalent to the theorem to be proved.

THEOREM 4 (G. CANTOR - H. LEBESGUE). *If at all points of a set E of positive measure,*

$$\lim_{n\to\infty}(a_n\cos nx + b_n\sin nx) = 0,$$

then

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = 0.$$

Proof. Set

$$r_n = \sqrt{a_n^2 + b_n^2}.$$

Then for arbitrary n, there exists an angle θ_n for which

$$a_n\cos nx + b_n\sin nx = r_n\cos(nx+\theta_n).$$

Assuming r_n does not tend to 0, we can choose a sequence $n_1 < n_2 < n_3 < \ldots$ for which $r_{n_k} > \sigma > 0$. But for $x \in E$,

$$r_n\cos(nx+\theta_n) \to 0;$$

this means that for these x,

$$\cos(n_k x + \theta_{n_k}) \to 0.$$

It follows, by Lebesgue's theorem on passage to the limit under the integral sign, that

$$\lim_{n\to\infty}\int_E \cos^2(n_k x+\theta_{n_k})\,dx=0. \tag{3}$$

On the other hand,

$$\int_E \cos^2(nx+\theta_n)\,dx=\frac{1}{2}\int_E[1+\cos(2nx+2\theta_n)]\,dx=$$

$$=\frac{1}{2}\Big[mE+\cos 2\theta_n\int_E \cos 2nx\,dx-\sin 2\theta_n\int_E \sin 2nx\,dx\Big].$$

The integrals

$$\int_E \cos nx\,dx,\int_E \sin nx\,dx,$$

being Fourier coefficients of the characteristic function of the set E, tend to 0 (as a consequence of the Riemann-Lebesgue theorem). Hence,

$$\lim_{n\to\infty}\int_E \cos^2(nx+\theta_n)\,dx=\frac{1}{2}mE,$$

which contradicts (3). This completes the proof of the theorem.

COROLLARY. *If a trigonometric series converges on a set of positive measure, then its coefficients tend to zero.*

The idea of the proof of the following theorem is very closely related to the above discussion.

THEOREM 5 (N. N. LUZIN - A. DENJOY). *If the trigonometric series*

$$\frac{a_0}{2}+\sum_{n=1}^{\infty}(a_n\cos nx+b_n\sin nx)$$

converges absolutely on a set E of positive measure, then

$$\sum_{n=1}^{\infty}(|a_n|+|b_n|)<+\infty.$$

Proof. Retaining the notation of the preceding theorem, we have, for $x\in E$, that

$$\sum_{n=1}^{\infty} r_n\cos^2(nx+\theta_n)\leqslant\sum_{n=1}^{\infty} r_n|\cos(nx+\theta_n)|<+\infty.$$

Let

$$\sum_{n=1}^{\infty} r_n\cos^2(nx+\theta_n)=A(x).$$

This function is measurable and finite on the set E. This means a subset E_0 ($mE_0>0$) can be chosen in E on which $A(x)$ is bounded and hence summable.

Then

$$\sum_{n=1}^{\infty} r_n \int_{E_0} \cos^2(nx+\theta_n)\,dx = \int_{E_0} A(x)\,dx < +\infty.$$

But in the proof of Theorem 4, we saw that

$$\int_{E_0} \cos^2(nx+\theta_n)\,dx \to \frac{mE_0}{2}.$$

This means that for $n \geqslant n_0$ this integral will be greater than $\frac{mE_0}{3}$ and, hence,

$$\frac{mE_0}{3}\sum_{n=n_0}^{\infty} r_n < \sum_{n=n_0}^{\infty} r_n \int_{E_0} \cos^2(nx+\theta_n)\,dx < +\infty,$$

from which the convergence of the series

$$\sum_{n=1}^{\infty} r_n$$

follows. This completes the proof of the theorem.

§5. SCHWARZ DERIVATIVES AND CONVEX FUNCTIONS

Our aim now is to discuss certain facts related to the questions of the uniqueness of the expansion of a function in a trigonometric series. To this end, we shall need certain new information which is of interest in itself. The present section will be devoted to the discussion of these ideas.

DEFINITION 1. Suppose the function $F(x)$ is defined in some neighborhood of the point x. If the limit

$$\lim_{h\to 0} \frac{F(x+h)-F(x-h)}{2h}$$

exists, it is called the *Schwarz derivative* of the function $F(x)$ at the point x and is denoted by $F^{(')}(x)$.

If the ordinary derivative $F'(x)$ exists at the point x, then the Schwarz derivative also exists and

$$F^{(')}(x) = F'(x).$$

This assertion follows directly from the fact that the right member of the equality

$$\frac{F(x+h)-F(x-h)}{2h} = \frac{1}{2}\left[\frac{F(x+h)-F(x)}{h} + \frac{F(x-h)-F(x)}{-h}\right]$$

tends to $F'(x)$ as $h \to 0$.

But situations occur where $F'(x)$ does not exist but $F^{(')}(x)$ exists. This is the case, for instance, at the point $x = 0$ for the function

$$F(x) = x \sin\frac{1}{x} \qquad [F(0) = 0].$$

Thus, the concept of the Schwarz derivative is a generalization of the concept of derivative. The notion of second derivative can be generalized in an analogous manner.

DEFINITION 2. Suppose the function $F(x)$ is defined in some neighborhood of the point x. If the limit

$$\lim_{h\to 0} \frac{F(x+h)-2F(x)+F(x-h)}{h^2}$$

exists, then it is called the *second Schwarz derivative* and is denoted by $F^{(n)}(x)$.

If the ordinary second derivative $F''(x)$ exists at the point x, then $F^{(n)}(x)$ also exists and

$$F^{(n)}(x) = F''(x).$$

In fact, the existence of the *second* derivative at the point x presupposes the existence of the *first* derivative $F'(x)$ in a neighborhood of this point. Setting

$$F(x+h)+F(x-h) = \varphi(h),$$

we can apply the well-known Cauchy formula to the right member of the equality

$$\frac{F(x+h)-2F(x)+F(x-h)}{h^2} = \frac{\varphi(h)-\varphi(0)}{h^2},$$

which yields

$$\frac{\varphi(h)-\varphi(0)}{h^2} = \frac{\varphi'(\theta h)}{2\theta h} \qquad (0<\theta<1),$$

and from which it follows that

$$\frac{F(x+h)-2F(x)+F(x-h)}{h^2} = \frac{F'(x+\theta h)-F'(x-h\theta)}{2\theta h}$$

and now the right member of this equality, as we have seen, tends to $F''(x)$ (as $h\to 0$).

$F^{(n)}(x)$ may exist even when $F''(x)$ does not exist; an example of such a function is

$$F(x) = \int_0^x t\sin\frac{1}{t}\,dt \tag{1}$$

for which $F^{(n)}(x)$ exists but $F''(x)$ does not exist at the point $x=0$.

THEOREM 1 (H. A. SCHWARZ). *Suppose the function $F(x)$ is defined and continuous on the closed interval $[a,b]$. If*

$$F^{(n)}(x) = 0$$

everywhere in the interval (a,b), then $F(x)$ is a linear function.

Proof. We take $\varepsilon>0$ and set

$$\varphi(x) = F(x) - \left[F(a) + \frac{F(b)-F(a)}{b-a}(x-a)\right] + \varepsilon(x-a)(x-b).$$

Clearly, $\varphi(x)$ is continuous on $[a,b]$ and $\varphi(a)=\varphi(b)=0$. Furthermore,

$$\varphi^{(n)}(x) = 2\varepsilon \tag{2}$$

everywhere on (a,b).

We shall show that

$$\varphi(x) \leqslant 0 \tag{3}$$

everywhere in $[a,b]$. In fact, if this were not the case, the function $\varphi(x)$ would assume its maximal value $\varphi(x_0)$ at some *interior* point x_0 and then the inequality

$$\frac{\varphi(x_0+h)-2\varphi(x_0)+\varphi(x_0-h)}{h^2}\leqslant 0$$

would, in the limit, yield

$$\varphi^{(n)}(x_0)\leqslant 0,$$

which contradicts (2).

In an analogous manner, we can verify that the function

$$\psi(x)=-\left\{F(x)-\left[F(a)+\frac{F(b)-F(a)}{b-a}(x-a)\right]\right\}+\varepsilon(x-a)(x-b)$$

satisfies the inequality

$$\psi(x)\leqslant 0 \tag{4}$$

everywhere.

Combining (3) and (4), we obtain

$$\left|F(x)-\left\{F(a)+\frac{F(b)-F(a)}{b-a}(x-a)\right\}\right|\leqslant\varepsilon|(x-a)(x-b)|,$$

from which it follows, since ε is arbitrary, that

$$F(x)=F(a)+\frac{F(b)-F(a)}{b-a}(x-a).$$

This completes the proof of the theorem.

We shall now take up the question of whether it is possible, knowing the second Schwarz derivative $F^{(n)}(x)$, to reconstruct the initial function $F(x)$. The course of ideas here reminds us very much of the line of reasoning we followed in §8, Chapter IX, when we considered the problem of the reconstruction of the primitive function starting with its derivative. We recommend that the reader refresh his memory by rereading that section at this point.

DEFINITION 3. The (finite or infinite) number λ will be called a *second Schwarz derived number* of the function $F(x)$ at the point x if there exists a sequence of positive numbers $h_1, h_2, h_3, \ldots$, tending to zero, such that

$$\lambda=\lim_{n\to\infty}\frac{F(x+h_n)-2F(x)+F(x-h_n)}{h_n^2}.$$

It is easily verified that any function $F(x)$ has a second Schwarz derived number at every point x and that a necessary and sufficient condition for the second Schwarz derivative to exist at the point x is that all the second Schwarz derived numbers be equal at this point.

DEFINITION 4. A function $F(x)$, defined on the closed interval $[a,b]$, is said to be *convex downward* if

$$F\left(\frac{x_1+x_2}{2}\right)\leqslant\frac{F(x_1)+F(x_2)}{2} \tag{5}$$

for any two points x_1 and x_2 of the closed interval $[a, b]$.

The simplest example of such a function is a linear function; in this case, the equality sign is attained in (5). Less trivial examples are supplied by the following lemma.

LEMMA 1. *If the function $f(t)$ increases on the closed interval $[a,b]$, then its indefinite integral*

$$F(x)=\int_a^x f(t)\,dt$$

is a function which is convex downward.

Proof. Let $a \leqslant x_1 < x_2 \leqslant b$. Then

$$F(x_1)-2F\left(\frac{x_1+x_2}{2}\right)+F(x_2)=\int_{\frac{x_1+x_2}{2}}^{x_2} f(t)\,dt-\int_{x_1}^{\frac{x_1+x_2}{2}} f(t)\,dt.$$

Since $f(t)$ increases, we have

$$\int_{\frac{x_1+x_2}{2}}^{x_2} f(t)\,dt \geqslant f\left(\frac{x_1+x_2}{2}\right)\frac{x_2-x_1}{2} \geqslant \int_{x_1}^{\frac{x_1+x_2}{2}} f(t)\,dt,$$

from which it follows that

$$F(x_1)-2F\left(\frac{x_1+x_2}{2}\right)+F(x_2)\geqslant 0.$$

Further, it is easily seen that

1) the sum of a finite number of functions which are convex downward is again a function which is convex downward;

2) the limit of a convergent sequence of functions which are convex downward is again a function which is convex downward;

3) the sum of a convergent series of functions which are convex downward is a function which is convex downward.

In going over to the differential properties of such functions, we first of all note that all the second Schwarz derived numbers of a function which is convex downward are *non-negative*. This property characterizes these functions. Namely, we have the following lemma.

LEMMA 2. *If all the second Schwarz derived numbers of a continuous function $F(x)$ are nonnegative, then $F(x)$ is convex downward.*

Proof. We take $\varepsilon > 0$ and consider, as in the proof of the Schwarz theorem, the function

$$\varphi(x)=F(x)-\left[F(a)+\frac{F(b)-F(a)}{b-a}(x-a)\right]+\varepsilon(x-a)(x-b).$$

$\varphi(x)$ is continuous; $\varphi(a)=\varphi(b)=0$; and all the second Schwarz derived numbers λ of this function satisfy the inequality $\lambda \geqslant 2\varepsilon$. It follows, as in the proof of the Schwarz theorem, that $\varphi(x)\leqslant 0$.

Passing to the limit as $\varepsilon \to 0$ in this inequality, we obtain

$$F(x)\leqslant F(a)+\frac{F(b)-F(a)}{b-a}(x-a)$$

and, in particular, that

$$F\left(\frac{a+b}{2}\right)\leqslant\frac{F(a)+F(b)}{2}.$$

This completes the proof of the lemma inasmuch as arbitrary numbers x_1 and x_2 could have been taken in place of a and b.

LEMMA 3. *Suppose the function $F(x)$ is defined and continuous on the closed interval $[a,b]$. If almost everywhere on (a,b) all the second Schwarz derived numbers are nonnegative and no second Schwarz derived number equals $-\infty$ at any point in (a,b), then the function $F(x)$ is convex downward.*

Proof. We denote the set of points in (a,b) at which at least one second Schwarz derived number of the function $F(x)$ is negative by E. By assumption, $mE=0$.

Therefore, in virtue of Theorem 6, §2, Chapt. VIII, there exists a continuous increasing function $\sigma(x)$ such that

$$\sigma'(x)=+\infty \tag{6}$$

at all points of the set E.

We set

$$\tau(x)=\int_a^x \sigma(t)\,dt.$$

By (6),

$$\tau''(x)=+\infty \tag{7}$$

at all points of the set E.

Having noted this, we now take an arbitrary $\varepsilon>0$ and consider the function

$$\Phi(x)=F(x)+\varepsilon\tau(x).$$

All the second Schwarz derived numbers of the function $\Phi(x)$ are nonnegative. In fact, by Lemma 1, the function $\tau(x)$ is convex downward, and therefore

$$\frac{\Phi(x+h)-2\Phi(x)+\Phi(x-h)}{h^2}\geqslant\frac{F(x+h)-2F(x)+F(x-h)}{h^2}.$$

Hence, if there existed a negative second Schwarz derived number of the function $\Phi(x)$ for $x\bar{\in}E$, then for the same x we should have that there exists a negative second Schwarz derived number also for $F(x)$, which is impossible by the very definition of the set E. But if $x\in E$, then the ratio

$$\frac{F(x+h)-2F(x)+F(x-h)}{h^2}$$

must be bounded below (otherwise a second Schwarz derived number would be $-\infty$ at this point) and therefore (in virtue of (7)),

$$\Phi^{(n)}(x)=+\infty.$$

Hence, $\Phi(x)$ satisfies the conditions of the preceding lemma and therefore it is convex downward. But then

$$F(x)=\lim_{\varepsilon\to 0}\Phi(x)$$

is also convex downward. This completes the proof of the lemma.

Now we can prove the result on the reconstruction of a function from its second Schwarz derivative which is the result we are interested in.

THEOREM 2 (C. J. DE LA VALLÉE POUSSIN). *Suppose $F(x)$ is a continuous function defined on $[a,b]$ and having a second Schwarz derivative $F^{(n)}(x)=f(x)$ everywhere in the interval (a,b).*

If $f(x)$ is everywhere finite and summable, then the equality

$$F(x)=\int_a^x dt\int_a^t f(u)\,du+Ax+B \tag{8}$$

is valid.

Proof. We now consider the function $\varphi_n(x)$, defined by setting

$$\varphi_n(x)=\begin{cases} f(x), \text{ if } f(x)\leqslant n,\\ n, \text{ if } f(x)>n.\end{cases}$$

Since

$$|\varphi_n(x)|\leqslant|f(x)|, \tag{9}$$

$\varphi_n(x)$ is summable.

Let

$$\Phi_n(x)=\int_a^x dt\int_a^t \varphi_n(u)\,du \text{ and } R_n(x)=F(x)-\Phi_n(x).$$

Since the integral $\int_a^t \varphi_n(u)\,du$ is continuous, we have

$$\Phi'_n(x)=\int_a^x \varphi_n(u)\,du$$

for all $x\in[a,b]$.

Now, since the right member of this equality has a derivative almost everywhere, equal to $\varphi_n(x)$, we have

$$\Phi''_n(x)=\varphi_n(x) \tag{10}$$

for almost all x. But everywhere where (10) holds, the difference $R_n(x)$ has a second Schwarz derivative,

$$R_n^{(n)}(x)=f(x)-\varphi_n(x)\geqslant 0.$$

Thus, the set of points where at least one second Schwarz derived number of the function $R_n(x)$ is negative has measure 0.

On the other hand, by the Cauchy formula,

$$\frac{\Phi_n(x+h)-2\Phi_n(x)+\Phi_n(x-h)}{h^2}=\frac{\Phi'_n(x+\theta h)-\Phi'_n(x-\theta h)}{2\theta h}=\frac{1}{2\theta h}\int_{x-\theta h}^{x+\theta h}\varphi_n(u)\,du\leqslant n$$

and hence

$$\frac{R_n(x+h)-2R_n(x)+R_n(x-h)}{h^2}\geqslant\frac{F(x+h)-2F(x)+F(x-h)}{h^2}-n.$$

From this it is clear that none of the second Schwarz derived numbers of the function $R_n(x)$ equals $-\infty$. Therefore, in virtue of Lemma 3, the function $R_n(x)$ is convex downward.

$\varphi_n(x)\to f(x)$ as $n\to\infty$. From this and (9) it follows that

$$\int_a^t \varphi_n(u)\,du\to\int_a^t f(u)\,du$$

for arbitrary t. But

$$\left| \int_a^t \varphi_n(u)\,du \right| \leqslant \int_a^b |f(u)|\,du$$

and therefore, for every x, we have

$$\Phi_n(x) = \int_a^x dt \int_a^t \varphi_n(u)\,du \to \int_a^x dt \int_a^t f(u)\,du.$$

Hence, the function

$$R(x) = F(x) - \int_a^x dt \int_a^t f(u)\,du,$$

being the limit function of convex functions $R_n(x)$, is itself convex downward.

This means that for arbitrary x_1 and x_2 we have

$$R\left(\frac{x_1+x_2}{2}\right) \leqslant \frac{R(x_1)+R(x_2)}{2}. \tag{11}$$

But, if we replace $F(x)$ by the function $-F(x)$, then we should have ended up in the above line of reasoning with the function $-f(x)$ instead of $f(x)$. Now the sign of $R(x)$ would change with the sign of $f(x)$ so that inequality (11) would assume the form

$$R\left(\frac{x_1+x_2}{2}\right) \geqslant \frac{R(x_1)+R(x_2)}{2}. \tag{12}$$

It follows from (11) and (12) that

$$R\left(\frac{x_1+x_2}{2}\right) = \frac{R(x_1)+R(x_2)}{2}$$

for arbitrary x_1 and x_2. This implies that for arbitrary $x \in (a,b)$ and sufficiently small $h > 0$, we shall have

$$R(x+h) - 2R(x) + R(x-h) = 0 \text{ and, consequently, } R^{(n)}(x) = 0.$$

By the Schwarz theorem, this implies that

$$R(x) = Ax + B,$$

and this is equivalent to the theorem we wished to prove.

The ideas discussed above suffice for the theory of trigonometric series. However, we wish to establish some more propositions on convex functions, in view of their inherent interest, although we shall not have occasion to refer to them later.

THEOREM 3. *If $f(x)$ is defined, bounded, and convex downward on the closed interval $[a,b]$, then it is continuous at all points of the interval (a,b).*

Proof. Suppose $a < x_0 < b$, and that $M(x)$ and $m(x)$ are the upper and lower Baire functions respectively of the function $f(x)$ (cf. §4, Chapt. V). We shall show that it is possible to find a sequence $\{x_n\}$ such that $x_n \to x_0$ and $f(x_n) \to M(x_0)$. In fact, if $n \to \infty$, then $M_{\frac{1}{n}}(x_0) \to M(x_0)$, and by the very definition of $M_{\frac{1}{n}}(x)$, there exists a point x_n in the interval $\left(x_0 - \frac{1}{n}, x_0 + \frac{1}{n}\right)$ such that $M_{\frac{1}{n}}(x) - \frac{1}{n} < f(x_n) \leqslant M_{\frac{1}{n}}(x_0)$. Evidently, $\{x_n\}$ is the required sequence.[15]

[15]One can, of course, also find a sequence $\{x_n\}$ such that $x_n \to x_0$, $f(x_n) \to m(x_0)$ analogously.

Having chosen such a sequence, we set

$$y_n = 2x_n - x_0.$$

Then also $y_n \to x_0$ and

$$f(x_n) = f\left(\frac{x_0 + y_n}{2}\right) \leqslant \frac{f(x_0) + f(y_n)}{2}.$$

We take an arbitrary $\varepsilon > 0$. If δ is sufficiently small, then[16]

$$M_\delta(x_0) < M(x_0) + \varepsilon,$$

and since $y_n \in (x_0 - \delta, x_0 + \delta)$ for sufficiently large n, we also have

$$f(x_n) < \frac{f(x_0) + M(x_0) + \varepsilon}{2}$$

for these n.

Passing to the limit, first as $n \to \infty$ and then as $\varepsilon \to 0$, we obtain

$M(x_0) \leqslant \frac{f(x_0) + M(x_0)}{2}$ from which it follows that $M(x_0) \leqslant f(x_0)$

and hence

$$M(x_0) = f(x_0). \tag{13}$$

This equality is also valid when $x_0 = a$ or $x_0 = b$ inasmuch as the point y_n lies on the same side of the point x_0 as the x_n.

We shall now find a sequence $\{x_n\}$ such that we have

$$x_n \to x_0,\ f(x_n) \to m(x_0),$$

and set

$$y_n = 2x_0 - x_n.$$

If $a < x_0 < b$, then the point y_n lies in $[a, b]$ for sufficiently large n (because $y_n \to x_0$) and then we have

$$f(x_0) = f\left(\frac{x_n + y_n}{2}\right) \leqslant \frac{f(x_n) + f(y_n)}{2}.$$

As above, for sufficiently large n, we shall have

$$f(x_0) \leqslant \frac{f(x_n) + M(x_0) + \varepsilon}{2},$$

from which it follows, by first passing to the limit as $n \to \infty$ and then as $\varepsilon \to 0$, that

$$f(x_0) \leqslant \frac{m(x_0) + M(x_0)}{2}.$$

In virtue of (13) and the evident relation $m(x) \leqslant M(x)$, it follows from this that $M(x_0) = m(x_0)$.

Now, according to the Baire theorem (§4, Chapt. V), $f(x)$ is continuous at the point x_0.

Remarks. 1. The condition that the function $f(x)$ be bounded is essential. There exist finite, everywhere discontinuous, convex functions which are not bounded in any interval.

---

[16] Since $f(x)$ is bounded, both Baire functions $M(x)$ and $m(x)$ are everywhere finite.

2. The function

$$f(x)=\begin{cases}0 \text{ if } -1<x<+1\\ 1 \text{ if } x=\pm 1\end{cases}$$

shows that a convex function may be discontinuous at the endpoints of its interval of definition.

THEOREM 4. *If $f(x)$ is convex downward, then*

$$f\left(\frac{x_1+x_2+\ldots+x_n}{n}\right)\leqslant\frac{f(x_1)+f(x_2)+\ldots+f(x_n)}{n} \tag{14}$$

for every natural number n.

Proof. If $n=2$, then (14) is equivalent to relation (5) in the definition of a function which is convex downward. Assume inequality (14) has been proved for $n=2^m$ and then let $n=2^{m+1}$. We set

$$x'=\frac{x_1+\ldots+x_{2^m}}{2^m},\quad x''=\frac{x_{2^m+1}+\ldots+x_{2^{m+1}}}{2^m}.$$

Then

$$f\left(\frac{x'+\ldots+x_n}{n}\right)=f\left(\frac{x'+x''}{2}\right)\leqslant\frac{f(x')+f(x'')}{2}\leqslant\frac{f(x_1)+\ldots+f(x_n)}{n}.$$

This proves (14) for all n of the form 2^m.

Now assume n is a number not of the form 2^m. We choose an m so large that $2^m>n$ and then set

$$A=\frac{x_1+x_2+\ldots+x_n}{n}.$$

Then

$$A=\frac{(x_1+\ldots+x_n)+(2^m-n)A}{2^m}$$

and, by what we already proved,

$$f(A)\leqslant\frac{f(x_1)+\ldots+f(x_n)+(2^m-n)f(A)}{2^m},\ \text{or}\ f(A)\leqslant\frac{f(x_1)+\ldots+f(x_n)}{n},$$

which completes the proof of the theorem. This ingenious line of reasoning is due to Cauchy.

COROLLARY. *If $p_1,p_2,\ldots,p_n$ are nonnegative numbers such that*

$$p_1+p_2+\ldots+p_n>0$$

and $f(x)$ is a continuous function, convex downward, then[17]

$$f\left(\frac{p_1x_1+\ldots+p_nx_n}{p_1+\ldots+p_n}\right)\leqslant\frac{p_1f(x_1)+\ldots+p_nf(x_n)}{p_1+\ldots+p_n}. \tag{15}$$

In fact, if all the p_i are rational, then (15) reduces to (14). The general case is obtained by passing to the limit [which requires that $f(x)$ be continuous], starting with rational p_i. In particular,

$$f(\alpha x+\beta y)\leqslant\alpha f(x)+\beta f(y)\qquad(\alpha\geqslant 0,\beta\geqslant 0,\alpha+\beta=1). \tag{16}$$

[17]Relation (15) is called the *Jensen summation inequality*.

Inequality (16) implies the following theorem.

THEOREM 5. *If $\Phi(u)$ is a continuous function, convex downward, defined on the entire real axis, then there exists a linear function $Au+B$ such that for all real u we have*

$$\Phi(u) > Au + B. \tag{17}$$

Proof. Suppose

$$l(u) = \frac{\Phi(1)-\Phi(-1)}{2}u + \frac{\Phi(1)+\Phi(-1)}{2}.$$

Clearly,

$$l(1) = \Phi(1), \quad l(-1) = \Phi(-1).$$

We shall show that

$$\Phi(u) \geqslant l(u) \tag{18}$$

for $|u| > 1$. Suppose, for instance, that $u > 1$. We set

$$\alpha = \frac{u-1}{u+1}, \qquad \beta = \frac{2}{u+1}.$$

If we set $x = -1, y = u$ in inequality (16), then it assumes the form

$$\Phi(1) \leqslant \frac{u-1}{u+1}\Phi(-1) + \frac{2}{u+1}\Phi(u),$$

from which (18) follows. The reasoning is analogous in case $u < -1$.

Since the function $\Phi(u)$ is continuous on the closed interval $[-1, +1]$ it is bounded and therefore a constant K can be found so large that for all $u \in [-1, +1]$ we have

$$K > \frac{\Phi(1)-\Phi(-1)}{2}u - \Phi(u). \tag{19}$$

(18) and (19) imply (17) if we take

$$A = \frac{\Phi(1)-\Phi(-1)}{2}, \quad B < -K, \quad B < \frac{\Phi(1)+\Phi(-1)}{2}.$$

COROLLARY. *Suppose $\Phi(u)$ satisfies the conditions of Theorem 5. If $f(x)$ and $p(x) \geqslant 0$ are defined on $[a,b]$, $f(x)$ is measurable and finite almost everywhere, and $p(x), p(x)f(x)$ are summable, then the integral*

$$\int_a^b \Phi[f(x)]p(x)\,dx \tag{20}$$

has a numerical value, either finite or equal to $+\infty$.

Proof. Without loss of generality, we may assume that $f(x)$ is finite. The composite function $\varphi(x) = \Phi[f(x)]$ is measurable inasmuch as it is for almost all x the limit function of a sequence of continuous functions $\Phi[f_n(x)]$, where $f_n(x)$ are continuous functions which satisfy the relation $f_n(x) \to f(x)$ almost everywhere. Suppose

$$\varphi_+(x) = \begin{cases} \varphi(x) \text{ for } \varphi(x) \geqslant 0, \\ 0 \quad \text{for } \varphi(x) < 0, \end{cases} \qquad \varphi_-(x) = \begin{cases} 0 \quad \text{for } \varphi(x) \geqslant 0, \\ -\varphi(x) \text{ for } \varphi(x) < 0. \end{cases}$$

Integral (20) has a numerical value whenever the difference

$$\int_a^b \varphi_+(x)p(x)\,dx - \int_a^b \varphi_-(x)p(x)\,dx$$

has a numerical value.

But according to Theorem 5, we have $\varphi(x) > Af(x) + B$. In particular, this inequality is satisfied on the set N of those x for which $\varphi(x) < 0$. Hence

$$0 \leqslant \varphi_-(x) \leqslant -Af(x) - B$$

for these x.

Multiplying this inequality by $p(x)$, we easily verify that the function $\varphi_-(x)p(x)$ is summable on N. But in the exterior of N, this product vanishes. Hence,

$$\int_a^b \varphi_-(x)p(x)\,dx \neq \infty.$$

The remainder of the proof is clear.

THEOREM 6. *If the conditions of the preceding corollary hold and*

$$\int_a^b p(x)\,dx > 0,$$

then the inequality

$$\Phi\left[\frac{\int_a^b f(x)p(x)\,dx}{\int_a^b p(x)\,dx}\right] \leqslant \frac{\int_a^b \Phi[f(x)]p(x)\,dx}{\int_a^b p(x)\,dx} \tag{21}$$

is valid; this is the so-called Jensen integral inequality.

Proof. Suppose, first, that both functions $f(x)$ and $p(x)$ are continuous. We set

$$x_k = a + \frac{b-a}{n}k \qquad (k = 0, 1, \ldots, n).$$

In virtue of (15), we have[18]

$$\Phi\left[\frac{f(x_0)p(x_0) + \ldots + f(x_{n-1})p(x_{n-1})}{p(x_0) + \ldots + p(x_{n-1})}\right] \leqslant \frac{\Phi[f(x_0)]p(x_0) + \ldots + \Phi[f(x_{n-1})]p(x_{n-1})}{p(x_0) + \ldots + p(x_{n-1})}.$$

It follows that

$$\Phi\left[\frac{\sum f(x_k)p(x_k)\Delta x_k}{\sum p(x_k)\Delta x_k}\right] \leqslant \frac{\sum \Phi[f(x_k)]p(x_k)\Delta x_k}{\sum p(x_k)\Delta x_k},$$

and (21) is obtained by passing to the limit as $n \to \infty$.

We shall now relinquish the assumption that $p(x)$ is continuous, while continuing for the time being to suppose that $f(x)$ is continuous.

It is easily verified that there exists a sequence of continuous functions $p_n(x) > 0$ for which

$$\lim_{n\to\infty} \int_a^b |p_n(x) - p(x)|\,dx = 0.$$

---

[18] It is easily verified that $p(x_0) + \ldots + p(x_{n-1}) > 0$ for large n.

By what we proved above, inequality (21) remains valid if we replace $p(x)$ by $p_n(x)$ in it. It now remains to pass to the limit as $n \to \infty$.

We shall now consider the case when $f(x)$ is measurable and bounded, $|f(x)| \leqslant K$. Then $\Phi[f(x)]$ is also a measurable and bounded function,

$$|\Phi[f(x)]| \leqslant M, \qquad M = \max_{-K \leqslant u \leqslant K} |\Phi(u)|.$$

Suppose the sequence of continuous functions $f_n(x)$ is such that

$$\lim_{n \to \infty} f_n(x) = f(x)$$

almost everywhere on $[a, b]$. In this connection, we may assume that $|f_n(x)| \leqslant K$. Then

$$|\Phi[f_n(x)]p(x)| \leqslant Mp(x), \qquad |f_n(x)p(x)| \leqslant Kp(x),$$

and we obtain (21) by first writing (21) with $f(x)$ replaced by $f_n(x)$ and then passing to the limit as $n \to \infty$.

Finally, we pass over to the general case. As we saw above, the right member of inequality (21) has a definite numerical value. We may assume this value is finite, for otherwise there would be nothing to prove.

But then, taking an arbitrary $\varepsilon > 0$, we can find a $\delta > 0$ such that $me < \delta$ implies the inequalities

$$\int_e |\Phi[f(x)]p(x)|\,dx < \varepsilon, \qquad \int_e p(x)\,dx < \varepsilon.$$

We construct a measurable bounded function $f_\varepsilon(x)$ for which

$$mE(f_\varepsilon \neq f) < \delta.$$

We may assume that $f_\varepsilon(x) = 0$ for those x at which $f_\varepsilon(x) \neq f(x)$. If $\varepsilon \to 0$, then $f_\varepsilon(x)$ converges in measure to $f(x)$. Furthermore, $|f_\varepsilon(x)| \leqslant |f(x)|$. Therefore,[19]

$$\lim_{\varepsilon \to 0} \int_a^b f_\varepsilon(x)p(x)\,dx = \int_a^b f(x)p(x)\,dx.$$

Hence, if we rewrite (21) with $f_\varepsilon(x)$ in place of $f(x)$, then the left member of the inequality thus obtained will tend to the left member of (21) as $\varepsilon \to 0$.

On the other hand,

$$\int_a^b \Phi(f)p\,dx - \int_a^b \Phi(f_\varepsilon)p\,dx = \int_{E(f_\varepsilon \neq f)} \Phi(f)p\,dx - \Phi(0) \int_{E(f_\varepsilon \neq f)} p\,dx.$$

It follows that

$$\left| \int_a^b \Phi(f)p\,dx - \int_a^b \Phi(f_\varepsilon)p\,dx \right| \leqslant \varepsilon\{1 + |\Phi(0)|\}.$$

Therefore, passing to the limit as $\varepsilon \to 0$ in inequality (21), modified as indicated above, we arrive at the right member of (21).

---

[19] Since $p(x)$ is finite almost everywhere, $f_\varepsilon p \Rightarrow fp$.

§6. UNIQUENESS OF THE TRIGONOMETRIC SERIES EXPANSION OF A FUNCTION

In this section we shall consider the question of the number of ways a given function can be expanded in a trigonometric series (provided such an expansion is at all possible).

We shall begin with the consideration of functions defined on the entire real line and having period 2π. Relative to such functions, the following elementary theorem is valid.

THEOREM 1. *If a function, defined on the entire real line, is expandable in a uniformly convergent trigonometric series, then this series is its Fourier series.*

This theorem can also be expressed in the following form.

THEOREM 1. *If a trigonometric series converges uniformly on the entire real line, then it is the Fourier series of its (obviously continuous) sum.*

The validity of this theorem follows from the fact that the equality

$$f(x)=\frac{a_0}{2}+\sum_{k=1}^{\infty}(a_k\cos kx+b_k\sin kx)$$

can be integrated term by term, first multiplying it by $\cos nx$ or $\sin nx$ (which, evidently, does not disturb the uniform convergence of the series). As a result, we obtain the equalities

$$a_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos nx\,dx, \qquad b_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin nx\,dx \quad (n=0,1,2,\dots),$$

which completes the proof of the theorem.

We shall need the following lemma in the sequel.

LEMMA 1 (B. RIEMANN). *Consider the convergent series*

$$a_0+a_1+a_2+\dots=S. \tag{1}$$

Starting with (1), *we construct the series*

$$a_0+a_1\left(\frac{\sin h}{h}\right)^2+a_2\left(\frac{\sin 2h}{2h}\right)^2+\dots\,. \tag{2}$$

Then series (2) *converges for arbitrary* $h\neq 0$ *and its sum* $S(h)$ *satisfies the relation*

$$\lim_{h\to 0}S(h)=S. \tag{3}$$

Proof. The convergence of series (1) implies that all its terms a_k are bounded by the same number, $|a_k|<M$, and therefore series (2) is majorized by the convergent series $\sum\frac{M}{k^2h^2}$ and, hence, converges.

We set

$$r_n=\sum_{k=n}^{\infty}a_k, \qquad r_n(h)=\sum_{k=n}^{\infty}a_k\left(\frac{\sin kh}{kh}\right)^2.$$

Taking $\varepsilon>0$, we can find an n such that

$$|r_k|<\varepsilon \tag{4}$$

for $k\geqslant n$.

We now keep this n fixed for the rest of our argument.

Since $a_k = r_k - r_{k+1}$, we have

$$r_n(h) = \sum_{k=n}^{\infty} (r_k - r_{k+1}) \left(\frac{\sin kh}{kh}\right)^2$$

from which it follows that

$$r_n(h) = r_n \left(\frac{\sin nh}{nh}\right)^2 + \sum_{k=n+1}^{\infty} r_k \left\{\left(\frac{\sin kh}{kh}\right)^2 - \left(\frac{\sin (k-1)h}{(k-1)h}\right)^2\right\}.$$

But

$$\left|\left(\frac{\sin kh}{kh}\right)^2 - \left(\frac{\sin (k-1)h}{(k-1)h}\right)^2\right| = \left|\int_{(k-1)h}^{kh} \frac{d}{dx}\left(\frac{\sin x}{x}\right)^2 dx\right| \leqslant \int_{(k-1)h}^{kh} \left|\frac{d}{dx}\left(\frac{\sin x}{x}\right)^2\right| dx.$$

Therefore,

$$|r_n(h)| \leqslant |r_n| + \sum_{k=n+1}^{\infty} |r_k| \int_{(k-1)h}^{kh} \left|\frac{d}{dx}\left(\frac{\sin x}{x}\right)^2\right| dx$$

and, in virtue of (4),

$$|r_n(h)| < \varepsilon \left\{1 + \int_{nh}^{+\infty} \left|\frac{d}{dx}\left(\frac{\sin x}{x}\right)^2\right| dx\right\}. \tag{5}$$

If we set[20]

$$L = \int_{0}^{+\infty} \left|\frac{d}{dx}\left(\frac{\sin x}{x}\right)^2\right| dx, \tag{6}$$

then inequality (5) will assume the form

$$|r_n(h)| < \varepsilon(1 + L).$$

Having noted this, we may, starting with the equality

$$S(h) - S = \sum_{k=1}^{n-1} a_k \left\{\left(\frac{\sin kh}{kh}\right)^2 - 1\right\} + r_n(h) - r_n,$$

assert that

$$|S(h) - S| \leqslant \sum_{k=1}^{n-1} |a_k| \cdot \left|\left(\frac{\sin kh}{kh}\right)^2 - 1\right| + (L+2)\,\varepsilon.$$

---

[20]This integral is finite. In fact

$$\frac{d}{dx}\left(\frac{\sin x}{x}\right)^2 = \frac{2\sin x}{x} \cdot \frac{x\cos x - \sin x}{x^2}.$$

This function is bounded for values of x near $x = 0$ and integral (6) is not improper; the inequality

$$\left|\frac{d}{dx}\left(\frac{\sin x}{x}\right)^2\right| \leqslant 2\frac{x+1}{x^3}$$

establishes its convergence also at infinity.

For fixed k, we have

$$\lim_{h\to 0}\left(\frac{\sin kh}{kh}\right)^2 = 1.$$

Therefore for given ε there corresponds a $\delta > 0$ such that

$$\sum_{k=1}^{n-1} |a_k| \cdot \left|\left(\frac{\sin kh}{kh}\right)^2 - 1\right| < \varepsilon$$

provided $|h| < \delta$.

For these h, we have

$$|S(h) - S| < (L+3)\,\varepsilon,$$

which completes the proof of the lemma.

We now consider the trigonometric series

$$\frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx), \tag{7}$$

subject to the condition that

$$|a_n| < M, \quad |b_n| < M \tag{8}$$

where M does not depend on n.

If we formally integrate series (7) twice, we obtain the series

$$\frac{a_0 x^2}{4} - \sum_{n=1}^{\infty}\frac{a_n \cos nx + b_n \sin nx}{n^2}. \tag{9}$$

In virtue of (8), series (9) is majorized by the series

$$\sum_{n=1}^{\infty}\frac{2M}{n^2}$$

and therefore it converges uniformly on the entire real line to some continuous function $F(x)$, which we shall call the *Riemann function* of series (7).

Thus, every series (7) satisfying condition (8), has a Riemann function. In particular, by the Cantor-Lebesgue theorem, §4, a series which converges on a set of positive measure, satisfies condition (8). We note that the convergence of series (7) in *one point* does not imply that it has a Riemann function. For example, the series

$$\sum_{n=1}^{\infty} n^2 \sin nx$$

which converges at the point $x = 0$, transforms, after integrating twice, into the series

$$-\sum_{n=1}^{\infty}\sin nx$$

which diverges[21] for $x \neq k\pi$.

---

[21] In fact, if $x \neq k\pi$, then

$$\sum_{k=1}^{n}\sin kx = \frac{\cos\frac{x}{2} - \cos\frac{2n+1}{2}x}{2\sin\frac{x}{2}},$$

and this quantity does not have a limit as $n \to \infty$.

DEFINITION. If the Riemann function $F(x)$ has a finite second Schwarz derivative at some point x_0, then this derivative $F^{(n)}(x)$ is called the *sum of series* (7) *at the point* x_0 *in the Riemann sense.*

THEOREM 2 (B. RIEMANN). *If series* (7) *satisfies condition* (8) *and converges at some point* x_0, *then it has a Riemann sum at this point which coincides with its ordinary sum.*

Proof. We start with the equality

$$F(x)=\frac{a_0x^2}{4}-\sum_{n=1}^{\infty}\frac{1}{n^2}(a_n\cos nx+b_n\sin nx);$$

after simple computations, we arrive at the equality

$$\frac{F(x_0+2h)-2F(x_0)+F(x_0-2h)}{4h^2}=$$

$$=\frac{a_0}{2}+\sum_{n=1}^{\infty}(a_n\cos nx_0+b_n\sin nx_0)\left(\frac{\sin nh}{nh}\right)^2. \tag{10}$$

Series (10) is obtained from the series

$$\frac{a_0}{2}+\sum_{n=1}^{\infty}(a_n\cos nx_0+b_n\sin nx_0) \tag{11}$$

in the same way that series (2) was obtained from (1). It follows, in virtue of Lemma 1, that

$$\lim_{h\to 0}\frac{F(x_0+2h)-2F(x_0)+F(x_0-2h)}{4h^2}=S,$$

where S is the sum of series (11). This completes the proof of the theorem.

This implies the following theorem.

THEOREM 3 (G. CANTOR). *If a trigonometric series converges on the entire real line and its sum is everywhere equal to zero, then all its coefficients are also equal to zero.*

Proof. According to the Cantor-Lebesgue theorem, the series satisfies condition (8) and hence it has a Riemann function $F(x)$. By the preceding theorem, the second Schwarz derivative $F^{(n)}(x)$ is everywhere equal to zero, $F^{(n)}(x)=0$, from which it follows, on the basis of the Schwarz theorem, §5, that the Riemann function $F(x)$ must be linear,

$$F(x)=Ax+B.$$

Thus,

$$Ax+B=\frac{a_0x^2}{4}-\sum_{n=1}^{\infty}\frac{a_n\cos nx+b_n\sin nx}{n^2} \tag{12}$$

for all real x.

Setting $x=\pi$ and then $x=-\pi$ in (12) and finally subtracting the equalities thus obtained, we can easily verify that $A=0$. In exactly the same way, by first setting $x=0$ and then $x=2\pi$ and finally subtracting, we obtain that $a_0=0$. Therefore, equality (12) assumes the form

$$-B=\sum_{n=1}^{\infty}\frac{a_n\cos nx+b_n\sin nx}{n^2}.$$

The series in the right member converges uniformly. Hence, by Theorem 1, it is the Fourier series of the left member, from which it follows that

$$\frac{a_n}{n^2}=\frac{1}{\pi}\int_{-\pi}^{\pi}(-B)\cos nx\,dx=0$$

and hence $a_n=0$. It is shown analogously that $b_n=0$. This completes the proof of the theorem.

THEOREM 4 (G. CANTOR). *If two everywhere convergent trigonometric series have the same sums, then they are identical, i.e. their corresponding coefficients are equal.*

In fact, subtracting these series termwise, we obtain a convergent expansion of zero.

All the coefficients of this expansion, according to what we have proved, must equal zero, and this implies the validity of Theorem 4. This theorem can also be formulated as follows: *the trigonometric series expansion of a function is unique.*

In connection with the theorem just proved, the question naturally arises of how the coefficients of a trigonometric series are determined by its sum. Theorem 1 gives the answer to this question in the case of uniform convergence. The following remarkable theorem is much more general.

THEOREM 5 (P. DU BOIS-REYMOND—C. J. DE LA VALLÉE POUSSIN). *If the sum of an everywhere convergent trigonometric series is (L)-integrable, then this series is its Fourier series.*

Proof. Suppose $f(x)$ is a finite summable function for which

$$f(x)=\frac{a_0}{2}+\sum_{n=1}^{\infty}(a_n\cos nx+b_n\sin nx) \tag{13}$$

for all x.

Since series (13) satisfies condition (8), it possesses a continuous Riemann function $F(x)$.

According to the Riemann theorem, $F^{(n)}(x)=f(x)$ everywhere, from which it follows, in virtue of the de la Vallée Poussin theorem, §5, that

$$F(x)=\Phi(x)+Ax+B,$$

where for brevity we have set

$$\Phi(x)=\int_0^x dt\int_0^t f(u)\,du.$$

Thus,

$$\Phi(x)=\frac{a_0x^2}{4}-Ax-B-\sum_{n=1}^{\infty}\frac{a_n\cos nx+b_n\sin nx}{n^2}.$$

After some minor calculations, it follows from the last equality that

$$\frac{\Phi(x+2h)-2\Phi(x)+\Phi(x-2h)}{4h^2}=\frac{a_0}{2}+\sum_{n=1}^{\infty}(a_n\cos nx+b_n\sin nx)\left(\frac{\sin nh}{nh}\right)^2.$$

The series in the right member converges uniformly and, consequently, it is the Fourier series of the left member. Hence,

$$a_n\left(\frac{\sin nh}{nh}\right)^2=\frac{1}{\pi}\int_0^{2\pi}\frac{\Phi(x+2h)-2\Phi(x)+\Phi(x-h)}{4h^2}\cos nx\,dx.$$

It follows that

$$\frac{4\pi a_n\sin^2 nh}{n^2}=\int_0^{2\pi}\varphi(x)\cos nx\,dx, \tag{14}$$

where for brevity we have set

$$\varphi(x)=\Phi(x+2h)-2\Phi(x)+\Phi(x-2h).$$

But if we now integrate by parts, we obtain

$$\alpha(h)=\int_0^{2\pi}\Phi(x+2h)\cos nx\,dx=\left[\Phi(x+2h)\frac{\sin nh}{n}\right]_0^{2\pi}-\int_0^{2\pi}\Phi'(x+2h)\frac{\sin nx}{n}\,dx$$

from which it follows that

$$\alpha(h)=-\frac{1}{n}\int_0^{2\pi}\left(\int_0^{x+2h}f(u)\,du\right)\sin nx\,dx.$$

Integrating by parts once more, we find that

$$\alpha(h)=\frac{1}{n^2}\int_{2h}^{2\pi+2h}f(u)\,du-\frac{1}{n^2}\int_0^{2\pi}f(x+2h)\cos nx\,dx.$$

Using the fact that $f(x)$ is periodic, this equality can be written in the form

$$\alpha(h)=\frac{1}{n^2}\int_0^{2\pi}f(x)\,[1-\cos n(x-2h)]\,dx.$$

It follows that

$$\int_0^{2\pi}\varphi(x)\cos nx\,dx=\alpha(h)-2\alpha(0)+\alpha(-h)=\frac{4\sin^2 nh}{n^2}\int_0^{2\pi}f(x)\cos nx\,dx.$$

Comparing this with (14), we obtain

$$a_n=\frac{1}{\pi}\int_0^{2\pi}f(x)\cos nx\,dx.$$

We verify in an analogous manner that a_0 and b_n are also Fourier coefficients of $f(x)$ and this completes the proof of the theorem.

Remarks. 1. Du Bois-Reymond proved this theorem for (R)-integrable functions and de la Vallée Poussin proved it for (L)-integrable functions.

2. Theorem 3 is a particular case of the Du Bois-Reymond-de la Vallée Poussin theorem. In fact, since the function which is identically equal to zero is summable, its trigonometric series expansion, being the Fourier series for zero, has all coefficients equal to zero.

3. In §4 we encountered everywhere convergent trigonometric series which were not Fourier series for any summable functions. It is clear that the sum of such a series is not (L)-integrable. In this connection, the problem arises of generalizing the Lebesgue integral in such a way that every everywhere convergent trigonometric series be the "Fourier series" of its sum. This problem was investigated by Denjoy.[22]

Up to the present, we considered *everywhere convergent* trigonometric series. For them, the uniqueness problem is solved by the Cantor and Du Bois-Reymond-de la Vallée Poussin theorems. We now consider the case of series which are not everywhere convergent. For these series, the uniqueness problem is formulated as follows.

It is known that *the trigonometric series*

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \tag{15}$$

converges to zero[23] *at all points of the closed interval* $[-\pi, \pi]$ *except, perhaps, at points of some set* $E \subset [-\pi, \pi]$.

If it follows from this proposition that all the coefficients of series (15) vanish,

$$a_n = b_n = 0 \tag{16}$$

(so that the series also converges to zero on the set E), then the set E is said to be a set of *type U* (from the French word unicité—uniqueness!). But if there exist series (15) which converge to zero in the exterior of E but which do not satisfy condition (16), then E is called a set of *type M* (from the French word multiplicité—multiplicity!). We are required to establish the conditions that the set E belong to one of these types.

The aforesaid problem was investigated by G. Cantor, W. Young, D. E. Menshov, N. K. Bari, A. Rajchman, and others. We shall give only the most elementary results in this area. This theme is discussed in significantly more detail in Zigmund: *Trigonometrical Series*, which we have already cited many times. The present state of this problem can be found in the survey article by N. K. Bari.[24]

THEOREM 6. *If the set A is a set of type M, then every set B containing it,* $B \supset A$, *is a set of type M.*

In fact, there exists a series with coefficients distinct from zero which converges to zero everywhere in the exterior of A. But then, a fortiori, it converges to zero everywhere in the exterior of B.

COROLLARY. *If B is a set of type U, then each of its subsets A contains a set of type U.*

---

[22]A. Denjoy, *Leçons sur le calcul des coefficients d'une série trigonométrique*, Paris, Gauthier-Villars, I, II, III (1941); IV_1, IV_2 (1949).

[23]We shall say that a series converges to zero if it converges and its sum equals zero.

[24]N. K. Bari, *The Problem of the Uniqueness of the Trigonometric Series Expansion of a Function*, Usp. Matem. Nauk, 4, No. 3 (31), 1949, pp. 3-68.

Roughly speaking, the more *opulent* the set E, the more probable it is that E is of type M; the more *meager* the set E, the greater the reason to believe it is of type U. The following theorem is a good illustration of what we have just stated.

THEOREM 7. *Every set E of positive inner measure $m_*E > 0$ is a set of type M.*

Proof. By the definition of inner measure, a closed set F of positive measure can be found, which is contained in E. We may assume that the points π and $-\pi$ do not lie in F. Then F is obtained from $[-\pi, \pi]$ by the removal of a finite number or of a denumerable set of pairwise disjoint intervals and the two intervals $[-\pi, \alpha)$, $(\beta, \pi]$ (where α and β are endpoints of F). Suppose $f(x)$ is the characteristic function of the set F. Then the Fourier series of this function converges to zero everywhere on the set $[-\pi, \pi] - F$ whereas as the same time,

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx = \frac{mF}{\pi} > 0,$$

and condition (16) is not satisfied.[25] Hence F and, a fortiori, $E \supset F$ are sets of type M.

Since we wish to give at least one example of a set of type U, we shall prove the following theorem.

THEOREM 8 (G. CANTOR). *Every finite set is a set of type U.*

We shall use the following four lemmas in the proof of this assertion.

LEMMA 2 (H. A. SCHWARZ). *Suppose the function $F(x)$ is continuous on the closed interval $[a, b]$. If*

$$F^{(n)}(x) = 0 \tag{17}$$

everywhere in (a, b), except perhaps for a finite number of points

$$x_1 < x_2 < \ldots < x_m \tag{18}$$

and

$$\lim_{h \to 0} \frac{F(x_k + h) - 2F(x_k) + F(x_k - h)}{h} = 0 \tag{19}$$

at the points (18), *then $F(x)$ is a linear function.*

Proof. According to the Schwarz theorem, §5, the function $F(x)$ must be linear on each of the closed intervals

$$[a, x_1], [x_1, x_2], \ldots, [x_m, b].$$

Let

$$F(x) = \begin{cases} Ax + B \text{ for } a \leqslant x \leqslant x_1, \\ Cx + D \text{ for } x_1 \leqslant x \leqslant x_2. \end{cases}$$

Condition (19), for the point x_1, yields

$$\lim_{h \to 0} \left[\frac{F(x_1 + h) - F(x_1)}{h} - \frac{F(x_1) - F(x_1 - h)}{h} \right] = 0,$$

from which it follows that $C = A$.

---

[25] The fact that (16) is not satisfied is also apparent from the Parseval equality

$$\frac{a_0^2}{2} + \sum_{n=1}^{\infty} (a_n^2 + b_n^2) = \frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x)\,dx = \frac{mF}{\pi} > 0.$$

On the other hand,

$$F(x_1) = Ax_1 + B = Cx_1 + D$$

and therefore $D = B$.

Thus, $F(x)$ is linear on the closed interval $[a, x_2]$. Its linearity on the entire closed interval $[a, b]$ is established in an analogous manner.

LEMMA 3. *If* $h > 0$, *then*

$$\sum_{k=1}^{\infty} \frac{\sin^2 kh}{k^2} < 3h. \tag{20}$$

Proof. We find a natural number n such that

$$n - 1 < \frac{1}{h} \leqslant n.$$

Then, in virtue of the inequality $|\sin\alpha| \leqslant |\alpha|$ we have

$$\sum_{k=1}^{n-1} \frac{\sin^2 kh}{k^2} \leqslant \sum_{k=1}^{n-1} \frac{k^2h^2}{k^2} = (n-1)h^2 < h.$$

On the other hand,[26]

$$\sum_{k=n}^{\infty} \frac{\sin^2 kh}{k^2} \leqslant \sum_{k=n}^{\infty} \frac{1}{k^2} < \frac{2}{n} \leqslant 2h,$$

which implies (20).

LEMMA 4. *If* $\lim_{n\to\infty} a_n = 0$, *then*

$$\lim_{h\to 0} \sum_{k=1}^{\infty} a_k \cdot \frac{\sin^2 kh}{k^2 h} = 0.$$

Proof. We take $\varepsilon > 0$ and find an n such that $|a_k| < \varepsilon$ for $k \geqslant n$.

Then, in virtue of (20),

$$\left| \sum_{k=n}^{\infty} a_k \frac{\sin^2 kh}{k^2 h} \right| < \frac{\varepsilon}{|h|} \sum_{k=1}^{\infty} \frac{\sin^2 kh}{k^2} < 3\varepsilon.$$

On the other hand, for fixed k, we have

$$\lim_{h\to 0} \frac{\sin^2 kh}{h} = 0$$

and, for sufficiently small $|h|$,

$$\left| \sum_{k=1}^{n-1} a_k \frac{\sin^2 kh}{k^2 h} \right| < \varepsilon.$$

For these h, we evidently have

---

[26] $$\sum_{k=n}^{\infty} \frac{1}{k^2} < \frac{1}{n^2} + \sum_{k=n+1}^{\infty} \frac{1}{k(k-1)} \leqslant \frac{1}{n} + \sum_{k=n+1}^{\infty} \left(\frac{1}{k-1} - \frac{1}{k} \right) = \frac{2}{n}.$$

$$\left|\sum_{k=1}^{\infty} a_k \frac{\sin^2 kh}{k^2 h}\right| < 4\epsilon,$$

which proves the lemma.

LEMMA 5. *If a trigonometric series satisfies the condition*

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = 0,$$

then, for any point x, the Riemann function $F(x)$ of this series satisfies the relation

$$\lim_{h\to 0} \frac{F(x+h) - 2F(x) + F(x-h)}{h} = 0.$$

Proof. Starting with the equality

$$F(x) = \frac{a_0 x^2}{4} - \sum_{n=1}^{\infty} \frac{a_n \cos nx + b_n \sin nx}{n^2}$$

which defines the Riemann function, we easily find that

$$\frac{F(x+2h) - 2F(x) + F(x-2h)}{4h} = \frac{a_0 h}{2} + \sum_{k=1}^{\infty} (a_n \cos nx + b_n \sin nx) \frac{\sin^2 nh}{n^2 h},$$

and everything reduces to the preceding lemma.

We can now give the

Proof of Theorem 8. Suppose a trigonometric series converges to zero everywhere in $[-\pi, \pi]$, except perhaps at the points $x_1 < x_2 < \ldots < x_m$. Then its coefficients tend to zero and, in virtue of Lemmas 5 and 2, and Theorem 2, the Riemann function of this series is linear in any finite interval and hence on the entire real line. After this remark, the proof is exactly the same as the proof of Theorem 3.

Theorem 8 was generalized by W. Young, who showed (1909) that every denumerable set is a set of type U. In connection with these results, the opinion was widely held by specialists that every set of measure 0 is a set of type U. However, in 1916, D. E. Menshov refuted this by constructing an example of a set of type M having measure 0. Of course, the set constructed by D. E. Menshov was nondenumerable. Then everyone suspected that in general all nondenumerable sets are of type M. But this opinion also turned out to be incorrect: in 1921, N. K. Bari and A. Rajchman, independently, constructed certain classes of perfect sets of type U. In particular, the Cantor perfect set P_0 turned out to be a set of type U. The problem of finding (non-tautological) conditions which are necessary and sufficient that a given set of measure 0 be of type U has not yet been solved.

EXERCISES FOR CHAPTER X

Let us agree to say that a point x is a *d-point* of the summable function $f(t)$ if the derivative of the indefinite integral of $f(t)$ equals $f(x)$ at this point (where $f(x) \neq \pm\infty$).

1. The integral

$$L_n(x)=\sqrt{\frac{n}{\pi}}\int_0^1[1-(t-x)^2]^n f(t)\,dt$$

is the Landau singular integral. Prove that if $x\ (0<x<1)$ is a d-point of the summable function $f(t)$, then $L_n(x)\to f(x)$. (F. Riesz)

2. The integral

$$P_r(x)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1-r^2}{1-2r\cos(t-x)+r^2}f(t)\,dt \qquad (0<r<1)$$

is the Poisson singular integral. Prove that if $x(-\pi<x<\pi)$ is a d-point of the summable function $f(t)$, then

$$\lim_{r\to 1-0}P_r(x)=f(x).$$

(P. Fatou)

3. The integral

$$V_n(x)=\frac{\sqrt{n}}{2\sqrt{\pi}}\int_{-\pi}^{\pi}\cos^{2n}\frac{t-x}{2}f(t)\,dt$$

is the de la Vallée Poussin singular integral. Prove: If $x(-\pi<x<\pi)$ is a d-point of the summable function $f(t)$, then $V_n(x)\to f(x)$. (C. J. de la Vallée Poussin)

4. The polynomial

$$K_n(x)=(n+1)\sum_{k=0}^{n}C_n^k x^k(1-x)^{n-k}\int_{\frac{k}{n+1}}^{\frac{k+1}{n+1}}f(t)\,dt$$

s the L. V. Kantorovič singular integral. Prove: If $x(0<x<1)$ is a d-point of the summable function $f(t)$, then $K_n(x)\to f(x)$. (L. V. Kantorovič)

5. Let $S_n(x)$ be the n-th partial sum of the Fourier series of the summable function $f(t)$ and let

$$B_n(x)=\frac{1}{2}\left[S_n(x)+S_n\left(x+\frac{2\pi}{2n+1}\right)\right].$$

$B_n(x)$ is the S. N. Bernstein singular integral. Prove: If $x(-\pi<x<\pi)$ is a Lebesgue point of $f(t)$, then $B_n(x)\to f(x)$. (I. P. Natanson)

6. Prove: A necessary condition that the equality

$$\lim_{n\to\infty}\int_a^b\varphi_n(t)f(t)\,dt=0$$

hold for every summable function $f(t)$ is that the functions $\varphi_n(t)$ be uniformly bounded, i.e. there exists a number K such that $|\varphi_n(t)|<K$ for all n. (H. Lebesgue)

7. Prove: It is impossible to construct a kernel $\Phi_n(t,x)$ so that the equality

$$\lim_{n\to\infty}\int_a^b \Phi_n(t,x)f(t)\,dt = f(x)$$

holds for an arbitrary summable function $f(t)$ which is approximately continuous at the point x. (I. P. Natanson)

8. Construct a kernel satisfying the conditions of the Lebesgue theorem, §2, but not satisfying the conditions of the Faddeyev theorem.

9. Suppose the kernel $\Phi_n(t,x)$ satisfies the conditions of the P. I. Romanovsky theorem. Let $F(t)$ be bounded and suppose the Stieltjes integral

$$I_n = \int_a^b \Phi_n(t,x)\,dF(t)$$

is meaningful for some $x \in (a,b)$. If $F'(x)$ exists and is finite for this x, then $I_n \to F'(x)$. (I. P. Natanson)

10. Derive P. I. Romanovsky's theorem from the result of Exercise 9.

11. Suppose $M(u)$ $(-\infty < u < +\infty)$ is an even function which is convex downward; $M(0) = 0$, $\lim\limits_{u\to+\infty} \dfrac{M(u)}{u} = +\infty$. A necessary and sufficient condition that

$$F(x) \quad (a \leqslant x \leqslant b)$$

be representable in the form

$$F(x) = C + \int_a^x f(t)\,dt,$$

where $\int_a^b M[f(t)]\,dt < +\infty$, is that the set of sums

$$\sum M\left(\frac{\Delta F(x_i)}{\Delta x_i}\right)\Delta x_i,$$

corresponding to arbitrary subdivisions of $[a,b]$, be bounded. (Yu. T. Medvedev)

12. A function which is convex downward satisfies a Lipschitz condition in every closed interval lying in the interior of the interval of definition of this function.

CHAPTER XI

POINT SETS IN TWO-DIMENSIONAL SPACE

§1. CLOSED SETS

Up to now we were concerned with the study of functions of one variable. For the theory of functions of several variables, we must know about point sets in multi-dimensional spaces. The present chapter is devoted to this topic. It should be noted that the most essential facts of the theory of multi-dimensional point sets can already be observed in the two-dimensional case, to which case we shall limit ourselves for the sake of simplicity of formulation and writing; the transition to spaces of a greater number of dimensions is met with only technical difficulties.

As in the linear case, the simplest sets are the closed and open sets. We begin with them.

DEFINITION 1. The point M_0 is said to be a *limit point* of the plane set E if every *open circle*, containing M_0, contains at least one point of the set E which is different from M_0.

Here, an open circle[1] is the set of points (x, y) whose coordinates satisfy the inequality

$$(x-a)^2+(y-b)^2<r^2,$$

where a, b and $r>0$ are constants. The point (a, b) is called the *center* of the circle and the number r its *radius*. If we adjoin to the open circle its boundary, i.e. those points for which $(x-a)^2+(y-b)^2=r^2$, we obtain a *closed* circle.

As in the linear case, it is easy to show that the point M_0 will be a limit point of the set E if, and only if, E contains a sequence of distinct points $M_1, M_2, M_3, \ldots$ such that

$$M_0=\lim M_n.$$

The last relation means the *distance* $\rho(M_0, M_n)$ tends to zero with increasing n. By the distance $\rho(A, B)$ between the points $A(a_1, a_2)$ and $B(b_1, b_2)$ we understand the number

$$\rho(A, B)=\sqrt{(b_1-a_1)^2+(b_2-a_2)^2}.$$

It is easy to show that if M_0 is a limit point of the set E, then every circle containing M_0 contains an infinite number of points of E.

The totality of all limit points of the set E is denoted by E' and is called the *derived* set of E. The points of the set $E-E'$ are isolated points of the set E.

In this connection, an important role is played by the following theorem.

THEOREM 1 (B. BOLZANO - K. WEIERSTRASS). *Every bounded infinite set E has at least one limit point.*

---

[1]It is assumed that for the plane case, the terms "circle," "center," "radius" are well-known to the reader. However, we stop to consider the purely arithmetic definitions of these concepts because we want to set an example which ought to be followed in going over to the n-dimensional case.

The proof is basically the same as in the linear case. Namely, the set E is enclosed in a rectangle $R(a \leqslant x \leqslant b, c \leqslant y \leqslant d)$ which is then subdivided into four rectangles by means of the straight lines

$$x = \frac{a+b}{2}, \quad y = \frac{c+d}{2};$$

of these rectangles, we choose one which contains an infinite number of points of the set E. Continuing this process, we obtain a sequence of nested rectangles, which contract to a point (x_0, y_0). It is easy to verify that (x_0, y_0) is a limit point of the set E.

It is not difficult to carry this theorem over to a sequence of points and obtain the following result.

THEOREM 2. *Every bounded sequence $M_1, M_2, M_3, \ldots$ of points in the plane contains a subsequence $M_{n_1}, M_{n_2}, M_{n_3}, \ldots (n_1 < n_2 < n_3 < \ldots)$ which converges to a point M_0,*

$$\lim M_{n_k} = M_0.$$

DEFINITION 2. A point set E is said to be *closed* if it contains all its limit points, i.e. if $E' \subset E$. If $E' = E$, the set E is said to be *perfect*.

Examples of closed sets are the closed circle $(x-a)^2 + (y-b)^2 \leqslant r^2$, the closed rectangle $a \leqslant x \leqslant b$, $c \leqslant y \leqslant d$, and also the set of points M whose coordinates x, y satisfy the inequality $F(x,y) \geqslant 0$, where $F(x,y)$ is an arbitrary continuous function defined on the whole plane.

We shall not study closed sets in as much detail as we did in the linear case: we shall limit ourselves to the consideration of only those properties of closed sets which will be directly used in the sequel.

THEOREM 3. *The intersection of an arbitrary collection of closed sets is a closed set. The sum of a finite number of closed sets is a closed set.*

The proof of this theorem does not differ at all from the linear case.

DEFINITION 3. Let E be a point set and let $\mathfrak{M}$ be a family of open circles. If for every point $M \in E$ there exists a circle $K \in \mathfrak{M}$ such that $M \in K$, then we say that the set E is *covered* by the family $\mathfrak{M}$.

THEOREM 4 (E. BOREL). *If the bounded closed set F is covered by an infinite family $\mathfrak{M}$ of open circles, then it is possible to select a finite subfamily $\mathfrak{M}^*$ from $\mathfrak{M}$ which also covers the set F.*

The proof, as in the linear case, is carried out by assuming the contrary. Namely, assuming that it is impossible to cover F with a finite number of circles of the family $\mathfrak{M}$, we enclose F in a rectangle $R(a \leqslant x \leqslant b, c \leqslant y \leqslant d)$ and subdivide it into four rectangles by the straight lines

$$x = \frac{a+b}{2}, \quad y = \frac{c+d}{2}.$$

At least one of them will contain a part of F which cannot be covered with a finite number of circles of the family $\mathfrak{M}$. Subdividing it into four parts and continuing this process, we arrive at a sequence of nested rectangles, each of which contains a part of F not covered by a finite subfamily of $\mathfrak{M}$.

These rectangles contract to a point M_0 which can easily be shown to belong to F. Then, as in the linear case, we obtain a contradiction without difficulty.

§2. OPEN SETS

DEFINITION 1. The point M_0 is called an *interior* point of the set E if there exists an open circle K which contains the point M_0 and is contained in the set E:

$$M_0 \in K \subset E.$$

DEFINITION 2. If every point of the set E is an interior point, then E is called an *open* set. An open circle can serve as an example of an open set.

THEOREM 1. *The sum of an arbitrary collection of open sets is an open set.*

The proof is the same as in the linear case.

THEOREM 2. *The intersection of a finite number of open sets is open.*

It will suffice to show that the intersection of any two open sets G_1 and G_2 is an open set. Let $M_0(x_0, y_0)$ be a point of this intersection. Then open circles

$$K_i: \quad (x-a_i)^2+(y-b_i)^2<r_i^2 \qquad (i=1,2)$$

can be found such that

$$M_0 \in K_i \subset G_i \qquad (i=1,2).$$

If we show that there exists a circle K which contains the point M_0 and is itself contained in the intersection of the circles K_1 and K_2, the theorem will be proved. It is not difficult to see that such a circle K is, say,

$$K: \qquad (x-x_0)^2+(y-y_0)^2<\rho^2,$$

where

$$\rho=\min_{i=1,2}\left\{r_i-\sqrt{(a_i-x_0)^2+(b_i-y_0)^2}\right\}.$$

We denote the complement of the plane set E with respect to the entire plane by CE. Then we can prove that the following theorem is valid in exactly the same way this was done in the linear case.

THEOREM 3. *The complement CG of an open set G is a closed set. The complement CF of a closed set F is an open set.*

Further, with the aid of the same reasoning as in §4, Chapt. II, we establish the following "separation theorem."

THEOREM 4. *If F_1 and F_2 are two disjoint bounded closed sets, then there exist two open sets G_1 and G_2 such that*

$$G_1 \supset F_1, \quad G_2 \supset F_2, \quad G_1 \cdot G_2 = 0.$$

The first question in which an essential difference between one-dimensional and two-dimensional open sets is found is that of the *structure* of open sets. Namely, in the two-dimensional case, we do not have the concept of *component interval* of an open set. Thus, the fundamental theorem on the structure of open sets for the plane case has a less clear-cut character than for the case of a linear set.

THEOREM 5. *Every non-void open set is the sum of a denumerable set of closed squares, the sides of which are parallel to the coordinate axes and which are pairwise without common interior points.*

Proof. Consider the two systems of straight lines

$$x=0, \pm 1, \pm 2, \pm 3, \ldots; y=0, \pm 1, \pm 2, \pm 3, \ldots .$$

These straight lines divide the entire plane into a denumerable number of squares (with each of which we include its boundary) whose sides have length 1. These squares are

pairwise without common interior points. We call them squares of the first rank. Next, we draw the straight lines

$$x = 0, \pm \tfrac{1}{2}, \pm 1, \pm \tfrac{3}{2}, \ldots; y = 0, \pm \tfrac{1}{2}, \pm 1, \pm \tfrac{3}{2}, \ldots .$$

The closed squares, into which the plane is subdivided by these straight lines, are called the squares of the second rank. It is clear that each square of the first rank consists of four squares of the second rank. Further, drawing the systems of lines

$$x = \frac{n}{4}, y = \frac{m}{4} \qquad (n, m = 0, \pm 1, \pm 2, \ldots)$$

$$x = \frac{n}{8}, y = \frac{m}{8} \qquad (n, m = 0, \pm 1, \pm 2, \ldots)$$

and continuing this process indefinitely, we obtain squares of rank 3, 4, 5, ...

All these squares are closed, their sides are parallel to the axes, any two squares of the same rank do not have common interior points, every square of rank k consists of four squares of rank $k+1$, and the side of a square of rank k has length 2^{1-k}; finally, the set of all such squares is denumerable.

Having made this preliminary construction, we now consider any non-void open set G. If M_0 is an arbitrary point in G, then it is possible to select (perhaps not in a unique manner) a sequence of nested squares of the first, second, third, ... ranks, containing the point M_0. Let these squares be

$$Q^{(1)} \supset Q^{(2)} \supset Q^{(3)} \supset \ldots .$$

Since the point M_0 is an interior point of the set G, and the length of the side of the square $Q^{(n)}$ tends to zero with increasing n, all squares $Q^{(n)}$, beginning with some one of them, are contained in G. Thus, there exist squares of the net constructed above which are contained in G. Having noted this, we denote by T_1 the set of all squares of the first rank contained in G, by T_2 the set of all those squares of the second rank which are contained in G but not contained in any of the squares of the collection T_1, by T_3 the set of all squares of the third rank which are contained in G but not contained in any square of the collections T_1 and T_2, and so on.

Each of the collections $T_1, T_2, T_3, \ldots$, and hence their sum T, is at most denumerable. Enumerate all squares of the collection T and let them be the squares $Q_1, Q_2, Q_3, \ldots$. We shall show that

$$G = \sum_{k=1}^{\infty} Q_k. \qquad (*)$$

To prove that G contains the sum of all Q_k is trivial. Conversely, if M_0 is a point of G, then, as we saw above, there exist squares of the initial net, contained in G and containing M_0. Let P be the set of numbers designating the ranks of such squares. P is a non-void set of positive integers; this means it contains a least integer. Let this least integer be m. Then there exists a square of rank m, containing M_0 and contained in G. This square is not contained in any square of the net which contains M_0, is contained in G, and has rank less than m, because in that case m would not be the least integer in P. But then the aforementioned square is in the collection T_m, which proves that

$$G \subset \sum_{k=1}^{\infty} Q_k.$$

Thus, equality (*) is proved. It remains to show that the right member has in fact a denumerable and not a finite number of terms. For if there were a finite number of terms, then G, being the sum of a finite number of closed squares, would be a closed set.

The only non-void plane set which is both open and closed is the entire plane,[2] but if G coincided with the entire plane, it of course would not be the sum of a finite number of squares, so that this last possibility is excluded.

It must be noted that the representation of an open set in the form indicated by this theorem is perhaps possible in more than one way; it is possible, for example, to subdivide only some of the squares considered, or at the very beginning to construct a net by dividing the squares not into 4 but into 9 parts, and so forth. This is also what we had in mind above when we said that in the plane case the fundamental theorem on the structure of an open set has a less clear-cut character than in the linear case.

To conclude this section, we shall prove the following generalization of Borel's theorem.

THEOREM 6. *Let $\mathfrak{M}$ be a family of open sets and let F be a bounded closed set. If every point of F belongs to at least one set $G \in \mathfrak{M}$, then $\mathfrak{M}$ contains a finite number of open sets whose sum contains F.*

In fact, every open set G is the sum of all the open circles contained in G. Therefore, F is covered by the collection of all open circles contained in the sets of the family $\mathfrak{M}$. Applying Borel's theorem to these circles, we obtain a finite number of them which cover F. Each of them is contained in some set of the family $\mathfrak{M}$. If we select these sets, we obtain the required finite subfamily of $\mathfrak{M}$.

From this it is clear, in particular, that in the Borel theorem we could have spoken of squares, for instance, instead of circles.

§3. THEORY OF MEASURE OF PLANE SETS

The theory of measure of plane sets is constructed exactly as in the linear case. There is a difference only in the initial points of the theory because a plane open set does not permit of a unique decomposition into component intervals and the complement of a bounded closed set with respect to the smallest rectangle containing it is not necessarily open.

DEFINITION 1. The area of an open rectangle $R(a<x<b, c<y<d)$, i.e. the number

$$mR=(b-a)(d-c),$$

is called the *measure* of the rectangle R.

The same number is called the measure of the closed rectangle $\bar{R}(a \leqslant x \leqslant b, c \leqslant y \leqslant d)$. (In both cases, it is assumed a priori that $a \leqslant b, c \leqslant d$. In the case where $a=b$, the open rectangle R is the void set.)

LEMMA 1. *Let a finite number of rectangles $R_1, R_2, \ldots, R_n$, with sides parallel to the axes, be given in the plane. These rectangles may or may not intersect. Some of them may be open, others closed. Then there exists a finite family of open rectangles $\gamma_1, \gamma_2, \ldots, \gamma_N$, with sides parallel to the axes, which possesses the following properties:*

1. *The rectangles γ_k are mutually disjoint;*

[2] Theorem 2, §4, Chapt. II carries over to the two-dimensional case.

2. *If* $R_i\gamma_k \neq 0$, *then* $R_i \supset \gamma_k$;

3. *The area of each* R_i *equals the sum of the areas of those* γ_k *which have points in common with* R_i.

Proof. Let

$$R_i = R_i(a_i \leqslant x \leqslant b_i, c_i \leqslant y \leqslant d_i),\ R_i = R_i(a_i < x < b_i, c_i < y < d_i),$$

(depending on whether R_i is closed or open).

Consider the set of points $\{a_i\}+\{b_i\}$ and enumerate them so that they form an increasing finite sequence

$$x_0 < x_1 < \ldots < x_p.$$

Analogously, let

$$y_0 < y_1 < \ldots < y_q$$

be the set $\{c_i\}+\{d_i\}$ arranged in increasing order. The required set of rectangles γ_k is obtained by considering the rectangles

$$(x_\lambda < x < x_{\lambda+1}, y_\mu < y < y_{\mu+1})$$
$$(\lambda = 0, 1, \ldots, p-1;\ \mu = 0, 1, \ldots, q-1).$$

The proof of this simple fact is left to the reader.

LEMMA 2. *Let* $\{\alpha_i\}$ *and* $\{\beta_j\}$ *be two finite families of rectangles,*[3] *where the rectangles* α_i *have, pairwise, no common interior points. If*

$$\sum \alpha_i \subset \sum \beta_j, \text{ then } \sum m\alpha_i \leqslant \sum m\beta_j.$$

To prove this, we combine both families $\{\alpha_i\}$ and $\{\beta_j\}$ into one common family $\{R_s\}$ to which we apply the preceding lemma. Since the rectangles α_i have no common interior points, no rectangle γ_k can be contained in two different α_i. Therefore, those γ_k which are contained in $\sum\alpha_i$ can be grouped into collections, where we put those γ_k which are contained in α_i into the collection T_i. It is important to note that the collections T_i are disjoint. Therefore

$$\sum m\alpha_i = \sum_i \Big(\sum_{T_i} m\gamma_k\Big) = \sum_T m\gamma_k,$$

where T is the set of all γ_k contained in $\sum\alpha_i$.

We now group the rectangles $\gamma_k \in T$ in another way. Namely, let S_1 be the set of those $\gamma_k \in T$ which are contained in β_1, S_2 the set of those $\gamma_k \in T$ which are contained in β_2 but not in S_1, S_3 the set of those $\gamma_k \in T$ which are contained in β_3 but not in $S_1 + S_2$, and so on.

Then

$$\sum_{S_j} m\gamma_k \leqslant m\beta_j,$$

because $m\beta_j$ equals the sum of the areas of all γ_k contained in β_j. This means that

$$\sum m\alpha_i = \sum_T m\gamma_k = \sum_j \Big(\sum_{S_j} m\gamma_k\Big) \leqslant \sum m\beta_j.$$

This completes the proof of the lemma.

COROLLARY. *If the rectangles of a finite or denumerable family* $\{\alpha_i\}$ *have, pairwise, no common interior points and all are contained in one rectangle* R, *then*

$$\sum m\alpha_i \leqslant mR.$$

---

[3] We have in mind rectangles whose sides are parallel to the axes; we shall not state this explicitly again because we shall not, in general, consider any other possibilities.

In fact, the case of a finite number of rectangles α_i is a particular case of the lemma, and the case of a denumerable set is obtained from this by passing to the limit.

LEMMA 3. *Let A and B be two plane bounded sets, each of which is representable in the form of the sum of a denumerable number of closed rectangles*

$$A=\sum_{i=1}^{\infty}\alpha_i, \quad B=\sum_{j=1}^{\infty}\beta_j,$$

where the rectangles α_i have, pairwise, no common interior points. If $A \subset B$, then

$$\sum_{i=1}^{\infty}m\alpha_i \leqslant \sum_{j=1}^{\infty}m\beta_j.$$

We note that the series in the right member may diverge because the rectangles β_j can overlap. But in that case, the lemma is trivial because the sum of a divergent series of positive terms equals $+\infty$. To prove the lemma under the condition that the series $\sum m\beta_j$ converges, we assume it is not true. Then it is possible to find an integer n so large that

$$\sum_{i=1}^{n}m\alpha_i > \sum_{j=1}^{\infty}m\beta_j. \tag{*}$$

Enclose each rectangle β_j in an *open* rectangle β_j^*, choosing β_j^* so that

$$m\beta^* < m\beta_j + \frac{\varepsilon}{2^j},$$

where ε is the positive difference between the left and right members of inequality (*). Then the family of open sets $\{\beta^*\}$ covers the bounded closed set

$$F=\sum_{i=1}^{n}\alpha_i.$$

According to the generalized Borel theorem, it is possible to find a finite number of rectangles $\beta_1^*, \beta_2^*, \ldots, \beta_m^*$ such that

$$F \subset \sum_{j=1}^{m}\beta^*.$$

By Lemma 2, we obtain

$$\sum_{i=1}^{n}m\alpha_i \leqslant \sum_{j=1}^{m}m\beta^* < \sum_{j=1}^{m}m\beta_j + \varepsilon,$$

from which, a fortiori,

$$\sum_{i=1}^{n}m\alpha_i < \sum_{j=1}^{\infty}m\beta_j + \varepsilon,$$

which contradicts the choice of the number ε.

In particular, the sets A and B may coincide. Therefore the following lemma is valid.

LEMMA 4. *If a plane bounded set E is representable as the sum of a denumerable number of closed rectangles in two ways:*

$$E=\sum_{i=1}^{\infty}\alpha_i=\sum_{j=1}^{\infty}\beta_j,$$

where the rectangles of each of these representations have, pairwise, no interior points in common, then

$$\sum_{i=1}^{\infty}m\alpha_i=\sum_{j=1}^{\infty}m\beta_j.$$

Every open set is the sum of a denumerable number of closed squares. It is natural to define the measure of such a set as the sum of the areas of these squares. This method is analogous to that accepted in the linear case where instead of squares we had component intervals. But here, in contrast to the linear case, we do not have *uniqueness* of the indicated decomposition and hence there could have arisen doubt as to whether we obtain the same sum of areas for different decompositions. Lemma 4 eliminates such doubts and we are justified in giving the following definition.

DEFINITION 2. If G is a plane bounded open set, the sum of the areas of the closed squares which have, pairwise, no common interior points and whose sum is G, is called the *measure* mG of G.

From Lemma 3 we have

THEOREM 1. *If G_1 and G_2 are bounded open sets and $G_1 \subset G_2$, then*

$$mG_1 \leqslant mG_2.$$

Further, the following theorem is valid.

THEOREM 2. *If a bounded open set G is the sum of a finite or denumerable number of mutually disjoint open sets G_k, then*

$$mG = \sum_k mG_k.$$

But if the condition that the sets G_k are mutually disjoint is omitted, then

$$mG \leqslant \sum_k mG_k.$$

In fact, let the equality

$$G_k = \sum_{i=1}^{\infty} \alpha_i^{(k)}$$

be a representation of G_k in the form of the sum of closed squares without common interior points. Then

$$G = \sum_{i,k} \alpha_i^{(k)}. \qquad (*)$$

If the sets G_k are pairwise disjoint, the individual squares $\alpha_i^{(k)}$ in representation (*) do not have, pairwise, interior points in common and by the very Definition 2, we have that

$$mG = \sum_{i,k} m\alpha_i^{(k)} = \sum_k \left(\sum_{i=1}^{\infty} m\alpha_i^{(k)}\right) = \sum_k mG_k.$$

But if we do not require the condition $G_k G_{k'} = 0$, $k \neq k'$, then G must be decomposed into squares by some other method, say,

$$G = \sum_{j=1}^{\infty} \delta_j$$

and, by Lemma 3,

$$mG = \sum_{j=1}^{\infty} m\delta_j \leqslant \sum_{i,k} m\alpha_i^{(k)} = \sum_k \left(\sum_{i=1}^{\infty} m\alpha_i^{(k)}\right) = \sum_k mG_k.$$

Thus, we see that although the definition of the measure of a closed set becomes somewhat more complicated in comparison with the linear case, its fundamental properties, expressed by the three theorems of §1, Chapt. III, are retained.

We now proceed to the definition of the measure of a bounded *closed* set F; we first enclose it in an *open* rectangle R. It is easy to see that the set

$$C_R F = R - F$$

is open; this is so because it can be written as the intersection $R \cdot CF$ of two open sets. We shall show that the number

$$mR - mC_R F \qquad (*)$$

does not depend on the choice of the rectangle R. In fact, if R_1 and R_2 are two such rectangles, then we can find a third rectangle R_3 containing both of them in its interior.

If we show that

$$mR_1 - mC_{R_1}F = mR_3 - mC_{R_3}F,\ mR_2 - mC_{R_2}F = mR_3 - mC_{R_3}F,$$

this will establish the requirement that the difference (*) is independent of R. It is thus sufficient to consider the case when $R_1 \subset R_2$.

We can assume that

$$R_1 = R_1(a_1 < x < b_1, c_1 < y < d_1),\ R_2 = R_2(a_2 < x < b_2, c_2 < y < d_2),$$

where

$$a_2 < a_1,\ b_2 > b_1,\ c_2 < c_1,\ d_2 > d_1.$$

This means that to go from R_1 to R_2, we must displace the left side to the left and the right side to the right in R_1, and the upper side upward, the lower side downward. These transformations can be carried out successively and since they are exactly of the same type, it is sufficient to compare the numbers

$$mR - mC_R F \quad \text{and} \quad mR' - mC_{R'}F,$$

where

$$R = R(a < x < b, c < y < d), \quad R' = R'(a < x < b + h, c < y < d)(h > 0).$$

But it is clear that

$$mR' = mR + h(d - c).$$

On the other hand,

$$C_{R'}F = (R' - R) + C_R F.$$

From this we can establish,[4] without difficulty, that

$$mC_{R'}F = h(d - c) + mC_R F,$$

which completes the proof of the assertion.

This enables us to introduce the following definition.

---

[4]This would be very simple if the set $R' - R$ were open, but this is not so. Therefore, the reasoning is somewhat more complicated. Namely, we take an $\varepsilon > 0$ and consider the open rectangles

$$U = U(b - \varepsilon < x < b + h, c < y < d),\ V = V(b + \varepsilon < x < b + h, c < y < d).$$

Obviously, $U \supset R' - R \supset V$. This means that

$$U + C_R F \supset C_{R'} F \supset V + C_R F$$

and

$$m[U + C_R F] \geqslant mC_{R'} F \geqslant m[V + C_R F].$$

The sets V and $C_R F$ do not intersect. Therefore

$$mU + mC_R F \geqslant mC_{R'} F \geqslant mV + mC_R F,$$

from which it follows that

$$(h + \varepsilon)(d - c) + mC_R F \geqslant mC_{R'} F \geqslant (h - \varepsilon)(d - c) + mC_R F,$$

and it remains to take the limit as $\varepsilon \to 0$.

DEFINITION 3. If F is a bounded closed set, the difference

$$mR - mC_R F,$$

where R is an arbitrary open rectangle containing the set F, is called the *measure* mF of F.

It is easy to show that if the set F is the sum of a finite number of closed squares, pairwise having no common interior points,

$$F = \sum_{k=1}^{n} \alpha_k,$$

then, according to the given definition, we have

$$mF = \sum_{k=1}^{n} m\alpha_k.$$

In fact, the set $C_R F$, being open, can be written in the form

$$C_R F = \sum_{i=1}^{\infty} \beta_i,$$

where the β_i are closed squares, pairwise having no common interior points. Then, a fortiori, α_k and β_i have no common points. This means that

$$R = \sum_{k=1}^{n} \alpha_k + \sum_{i=1}^{\infty} \beta_i \text{ and therefore } mR = \sum_{k=1}^{n} m\alpha_k + \sum_{i=1}^{\infty} m\beta_i,$$

from which it follows that

$$\sum_{k=1}^{n} m\alpha_k = mR - \sum_{i=1}^{\infty} m\beta_i = mR - mC_R F = mF.$$

If we return to §2, Chapt. III, and observe all the theorems on the measure of a bounded closed linear set, we notice that they are all based not so much on the definition of the measure mF as on the lemma stating that

$$mF = m\Delta - mC_\Delta F,$$

where Δ is an arbitrary interval containing F, and on the fact that the measure of a finite number of disjoint segments equals the sum of their lengths.[5] But we saw that both these properties carry over into the theory of plane sets. Therefore, without going into further detail, we can assert that the entire theory of measure for closed sets carries over from the linear to the plane case. So, for example, if $F_1 \subset F_2$, then $mF_1 \leqslant mF_2$; if G is a bounded open set, its measure is the least upper bound of the measures of the closed sets contained in it; and so on.

But then, exactly as in §3, Chapt. III, we can define the concept of outer and inner measure of an arbitrary bounded set E. All seven theorems of §3, Chapt. III, carry over to the two-dimensional case without modification, except that we speak of open rectangles instead of the intervals Δ, referred to in Chapt. III.

Finally, we can introduce the fundamental definition of a measurable set as a bounded set for which the inner measure equals the outer measure. This done, the context of §§4 and 6, Chapt. III is carried over to the plane case without difficulty. We shall not go into detail here and shall only extend the material already known by one proposition which will be useful in our further discussion.

---

[5] This property was used in Theorem 4, §2, Chapt. III.

THEOREM 3. *For every bounded set E, there exist sets A and B possessing the following properties*:

1) *A is a set of type* F_σ *and B is a set of type* G_δ;

2) $A \subset E \subset B$;

3) $mA = m_*E, mB = m^*E$.

For the proof, we assign to every natural number n a closed set F_n and an open set G_n such that

$$F_n \subset E \subset G_n,\ mF_n > m_*E - \frac{1}{n},\ mG_n < m^*E + \frac{1}{n};$$

then we set

$$A = F_1 + F_2 + F_3 + \ldots,\ B = G_1G_2G_3\ldots .$$

As far as the Vitali theorem, which was the subject matter of §8, Chapt. III, is concerned, it carries over to the plane case without change. Namely, let us agree to say that the set E is covered by a family $\mathfrak{M}$ of closed squares[6] *in the Vitali sense* if for every point x of E, squares of arbitrarily small area, containing x, can be found in the family $\mathfrak{M}$. With the aid of this definition, we can establish two forms of the Vitali theorem by repeating almost verbatim the line of reasoning of §8, Chapt. III.

THEOREM 4. *If a plane bounded set E is covered in the Vitali sense by a family of closed squares* $\mathfrak{M} = \{Q\}$, *then* $\mathfrak{M}$ *contains a denumerable set of pairwise disjoint squares* $Q_1, Q_2, Q_3, \ldots$, *whose sum covers the set E to within a set of measure zero, i.e.*

$$Q_kQ_{k'} = 0\ \ (k \neq k'), \qquad m\left[E - \sum_{k=1}^{\infty} Q_k\right] = 0.$$

THEOREM 5. *Under the conditions of the preceding theorem, to every* $\varepsilon > 0$ *there corresponds a finite set* $Q_1, Q_2, \ldots, Q_n$ *of pairwise disjoint squares in* $\mathfrak{M}$ *such that*

$$m^*\left[E - \sum_{k=1}^{n} Q_k\right] < \varepsilon.$$

The questions of the invariance of measure under isometries, treated in §5, Chapt. III, are of a special nature. Carrying over the corresponding results to the multi-dimensional case presents essential difficulties; we devote the next section to this topic.

§4. MEASURABILITY AND MEASURE AS INVARIANTS UNDER ISOMETRIES

DEFINITION 1. A single-valued mapping $M^* = \varphi(M)$ of the plane R_2 into itself is an *isometry* if the distance between the images of any two points in the plane equals the distance between these points themselves:

$$\rho[\varphi(M), \varphi(N)] = \rho(M, N).$$

It is clear that distinct images M^* correspond to distinct points M, so that an isometry is a transformation whose inverse is single-valued.

Let the points $P_0(\alpha_0, \beta_0)$, $P_1^*(\alpha_1, \beta_1)$ and $P_2^*(\alpha_2, \beta_2)$ be the images of the points $P_0(0, 0)$, $P_1(1, 0)$ and $P_2(0, 1)$.

Since $\rho(P_0, P_1) = 1, \rho(P_0, P_2) = 1, \rho(P_1, P_2) = \sqrt{2}$, the distances between their images are respectively the same. This means that

---

[6] Recall that we always assume the sides of the squares considered to be parallel to the axes.

$$(\alpha_1-\alpha_0)^2+(\beta_1-\beta_0)^2=1,\ (\alpha_2-\alpha_0)^2+(\beta_2-\beta_0)^2=1,$$
$$(\alpha_2-\alpha_1)^2+(\beta_2-\beta_1)^2=2.$$

Writing the differences $\alpha_2-\alpha_1$ and $\beta_2-\beta_1$ in the form

$$(\alpha_2-\alpha_0)-(\alpha_1-\alpha_0),\qquad (\beta_2-\beta_0)-(\beta_1-\beta_0),$$

we obtain without difficulty from the preceding equalities that

$$(\alpha_2-\alpha_0)(\alpha_1-\alpha_0)+(\beta_2-\beta_0)(\beta_1-\beta_0)=0.$$

Having noted this, we consider the determinant

$$D=\begin{vmatrix}\alpha_1-\alpha_0 & \beta_1-\beta_0\\ \alpha_2-\alpha_0 & \beta_2-\beta_0\end{vmatrix};$$

its square, formed by the rule of multiplying "row by row," equals

$$\begin{vmatrix}1 & 0\\ 0 & 1\end{vmatrix}=1,$$

and therefore $|D|=1$.

Further, let $M(x,y)$ be an arbitrary point in the plane, and let $M^*(x^*,y^*)$ be its image. Equating the distances $\rho(M,P_i)$ and $\rho(M^*,P_i^*)$ $(i=0,1,2)$, we obtain 3 equations:

$$x^2+y^2=(x^*-\alpha_0)^2+(y^*-\beta_0)^2,$$
$$(x-1)^2+y^2=(x^*-\alpha_1)^2+(y^*-\beta_1)^2,$$
$$x^2+(y-1)^2=(x^*-\alpha_2)^2+(y^*-\beta_2)^2.$$

Subtracting the second and third equations from the first, we find that

$$\left.\begin{aligned}x&=(\alpha_1-\alpha_0)x^*+(\beta_1-\beta_0)y^*+\gamma_1,\\ y&=(\alpha_1-\alpha_0)x^*+(\beta_2-\beta_0)y^*+\gamma_2,\end{aligned}\right\}\tag{1}$$

where γ_1 and γ_2 are constants whose exact expressions are not needed.

We shall show that every point $Q(\lambda,\mu)$ in the plane is the image of some point $M(x,y)$. In fact, let

$$\begin{aligned}x&=(\alpha_1-\alpha_0)\lambda+(\beta_1-\beta_0)\mu+\gamma_1,\\ y&=(\alpha_2-\alpha_0)\lambda+(\beta_2-\beta_0)\mu+\gamma_2.\end{aligned}\tag{2}$$

The image of the point $M(x,y)$ must be found in terms of x and y from equations (1). But the determinant of system (1) is $D\neq 0$. This means that there is *only one* pair of numbers (x^*,y^*), satisfying equations (1), and then it is clear from (2) that $x^*=\lambda, y^*=\mu$.

Thus, an isometry is a one-to-one mapping of the plane R_2 onto itself. It is clear that the inverse mapping is also an isometry. But then, interchanging the roles of the points, images and inverse images, we see that the coordinates of the image are expressed in terms of the coordinates of the inverse image by equations of type (1). We have thus proved the following theorem.

THEOREM 1. *The coordinates (x^*,y^*) of the image point M^* of the point $M(x,y)$, under an isometry, are expressed by the formulas*

$$x^*=a_1x+b_1y+c_1,\qquad y^*=a_2x+b_2y+c_2.\tag{3}$$

If

$$\Delta=\begin{vmatrix}a_1 & b_1\\ a_2 & b_2\end{vmatrix},\tag{4}$$

then $|\Delta|=1$.

DEFINITION 2. Any transformation $M^* = \varphi(M)$ of the plane R_2 into itself for which the coordinates of the image $M^*(x^*, y^*)$ are expressed in terms of the coordinates of the inverse image $M(x, y)$ by means of the formulas (3) is called an *affine* transformation. If in this connection determinant (4) (called the *determinant of the transformation*) is different from zero, then we say that the transformation is *non-degenerate.*

Thus, the following theorem is valid.

THEOREM 2. *Every isometry is an affine transformation the modulus of whose determinant equals one,*[7] $|\Delta| = 1$.

Further, the following theorem holds.

THEOREM 3. *Two affine transformations carried out in succession are equivalent to a single affine transformation whose determinant equals the product of the determinants of the initial transformations.*

The simple proof of this fact is left to the reader.

THEOREM 4. *Let* $R(\alpha \leqslant x \leqslant \beta, \gamma \leqslant y \leqslant \delta)$ *be a closed rectangle. As a result of each of the following four affine transformations*

$$\text{I)}\ \begin{matrix} x^* = y \\ y^* = x \end{matrix} \qquad \text{II)}\ \begin{matrix} x^* = x + y \\ y^* = y \end{matrix} \qquad \text{III)}\ \begin{matrix} x^* = x + a \\ y^* = y + b \end{matrix}$$

$$\text{IV)}\ \begin{matrix} x^* = kx \\ y^* = ly \end{matrix} \qquad (kl \neq 0),$$

the rectangle R goes over into the measurable set R, whose measure is*

$$mR^* = |\Delta| \cdot mR,$$

where Δ *is the determinant of the transformation.*

In fact, in cases I, III and IV the rectangle R goes over into the rectangle R^* defined, respectively, by the inequalities

$$\text{I)}\ \begin{matrix} \gamma \leqslant x^* \leqslant \delta \\ \alpha \leqslant y^* \leqslant \beta \end{matrix} \qquad \text{III)}\ \begin{matrix} \alpha + a \leqslant x^* \leqslant \beta + a \\ \gamma + b \leqslant \gamma^* \leqslant \delta + b \end{matrix} \qquad \text{IV)}\ \begin{matrix} k\alpha \leqslant x^* \leqslant k\beta \\ l\gamma \leqslant y^* \leqslant l\delta. \end{matrix}$$

(The last inequalities are written for the case $k > 0$, $l > 0$. If, for instance, $k < 0$, the first of them must be replaced by the inequality $k\beta \leqslant x^* \leqslant k\alpha$.)

From this the validity of the theorem for transformations I), III), IV) is clear. The study of transformation II) is significantly more complicated. We note, first of all, that in case II) R^* is defined by the inequalities

$$\alpha \leqslant x^* - y^* \leqslant \beta, \qquad \gamma \leqslant y^* \leqslant \delta.$$

Further, we take a natural number n and subdivide R^* into parts

$$R^* = S_1 + S_2 + \ldots + S_n,$$

where S_k, for $k = 1, 2, \ldots, n-1$, is defined by the inequalities

$$\alpha \leqslant x^* - y^* \leqslant \beta, \qquad \gamma + \frac{k-1}{n}(\delta - \gamma) \leqslant y^* < \gamma + \frac{k}{n}(\delta - \gamma),$$

and S_n is defined by the inequalities

---

[7] It is useful to note that the converse theorem is not true. For example, the transformation $x^* = 2x$, $y^* = \frac{1}{2}x$ is not an isometry. The affine transformation (3) will be an isometry if and only if

$$a_1^2 + a_2^2 = 1,\ b_1^2 + b_2^2 = 1,\ a_1 b_1 + a_2 b_2 = 0.$$

$$\alpha \leqslant x^* - y^* \leqslant \beta, \qquad \gamma + \frac{n-1}{n}(\delta - \gamma) \leqslant y^* \leqslant \delta.$$

It is evident that the sets S_k are pairwise disjoint.

It is easy to see that $U_k \subset S_k \subset V_k$, where U_k and V_k are rectangles defined by the inequalities

$$U_k = U_k\Big[\alpha + \gamma + \frac{k}{n}(\delta - \gamma) \leqslant x^* \leqslant \beta + \gamma + \frac{k-1}{n}(\delta - \gamma),$$
$$\gamma + \frac{k-1}{n}(\delta - \gamma) \leqslant y^* < \gamma + \frac{k}{n}(\delta - \gamma)\Big],$$

$$V_k = V_k\Big[\alpha + \gamma + \frac{k-1}{n}(\delta - \gamma) \leqslant x^* \leqslant \beta + \gamma + \frac{k}{n}(\delta - \gamma),$$
$$\gamma + \frac{k-1}{n}(\delta - \gamma) \leqslant y^* \leqslant \gamma + \frac{k}{n}(\delta - \gamma)\Big].$$

Therefore,

$$mU_k \leqslant m_* S_k \leqslant m^* S_k \leqslant mV_k$$

or

$$\Big(\beta - \alpha - \frac{\delta - \gamma}{n}\Big)\frac{\delta - \gamma}{n} \leqslant m_* S_k \leqslant m^* S_k \leqslant \Big(\beta - \alpha + \frac{\delta - \gamma}{n}\Big)\frac{\delta - \gamma}{n}.$$

Summing these inequalities and noting that

$$m_* R^* \geqslant \sum_{k=1}^{n} m_* S_k, \qquad m^* R^* \leqslant \sum_{k=1}^{n} m^* S_k,$$

we obtain

$$(\beta - \alpha)(\delta - \gamma) - \frac{(\delta - \gamma)^2}{n} \leqslant m_* R^* \leqslant m^* R^* \leqslant (\beta - \alpha)(\delta - \gamma) + \frac{(\delta - \gamma)^2}{n}.$$

Since n is arbitrary, it follows from this that

$$m_* R^* = m^* R^* = (\beta - \alpha)(\delta - \gamma) = mR.$$

This completes the proof of the theorem.

THEOREM 5. *Let G be a bounded open set. As a result of each of the affine transformations I, II, III, IV considered, the set G goes over into a measurable set G^*, whose measure is*

$$mG^* = |\Delta| \cdot mG,$$

where Δ is the determinant of the transformation.

In fact, the condition $\Delta \neq 0$ shows that every non-degenerate affine transformation $M^* = \varphi(M)$, being one-to-one, preserves incidence, i.e. the relations

$$M \in E, \quad A \subset B, \quad A \cdot B = 0, \quad E = \sum_{k=1}^{\infty} E_k$$

imply the relations

$$M^* \in E^*, \quad A^* \subset B^*, \quad A^* \cdot B^* = 0, \quad E^* = \sum_{k=1}^{\infty} E_k^*,$$

where the notation used is self-evident.

Furthermore, it is not difficult to see that as a result of an affine transformation, a bounded set goes over again into a bounded set.

Having noted this, we take any bounded open set G and write it in the form

$$G=\sum_{k=1}^{\infty} R_k,$$

where the R_k are closed squares without common interior points. If $G^*=\varphi(G)$ and $R_k^*=\varphi(R_k)$ are the images of the sets G and R_k under one of the transformations I — IV, then

$$G^*=\sum_{k=1}^{\infty} R_k^*.$$

From this the measurability of G^* and the validity[8] of the equalities

$$mG^*=\sum_{k=1}^{\infty} mR_k^*=\sum_{k=1}^{\infty}|\Delta|\cdot mR_k=|\Delta|\cdot mG$$

are clear.

This completes the proof of the theorem.

THEOREM 6. *Every non-degenerate affine transformation* (3) *is the result of a sequence of the simpler transformations I — IV.*

In fact, if in the transformation

$$x^*=a_1x+b_1y+c_1,\quad y^*=a_2x+b_2y+c_2$$

none of the coefficients a_1, a_2, b_1 is zero, then this transformation is the resultant of the following eight transformations of type I — IV:

$$\left|\begin{array}{l} x_1=a_1x \\ y_1=b_1y \end{array}\right| \quad \left|\begin{array}{l} x_2=x_1+y_1 \\ y_2=y_1 \end{array}\right| \quad \left|\begin{array}{l} x_3=y_2 \\ y_3=x_2 \end{array}\right| \quad \left|\begin{array}{l} x_4=\dfrac{\Delta}{a_1b_1}x_3 \\ y_4=\dfrac{a_2}{a_1}y_3 \end{array}\right|$$

$$\left|\begin{array}{l} x_5=x_4+y_4 \\ y_5=y_4 \end{array}\right| \quad \left|\begin{array}{l} x_6=x_5 \\ y_6=\dfrac{a_1}{a_2}y_5 \end{array}\right| \quad \left|\begin{array}{l} x_7=y_6 \\ y_7=x_6 \end{array}\right| \quad \left|\begin{array}{l} x^*=x_7+c_1 \\ y^*=y_7+c_2 \end{array}\right|.$$

In the case where certain of the indicated coefficients are zero, the reasoning only becomes simpler.

THEOREM 7. *As a result of a non-degenerate affine transformation, a bounded plane set E goes over into the bounded plane set E^*, whose outer measure is*

$$m^*E^*=|\Delta|\cdot m^*E,$$

where Δ is the determinant of the transformation.

In fact, let us first of all carry out one of the simple transformations of type I, II, III, IV. Taking an arbitrary bounded set E, we find a bounded open set G such that

$$G\supset E,\ mG<m^*E+\varepsilon,$$

---

[8]We must note that the sets R_k^* can intersect because the squares R_k can have common sides. But every such side is a rectilinear segment, parallel to one of the axes, and hence the measure of such a side equals zero. We can consider every segment parallel to an axis as being on a rectangle. This means that, according to the preceding theorem, the measure of the image of such a segment is 0. Therefore, the set of points common to several R_k^* has measure zero and, in finding the value of mG^*, such a set can be disregarded.

where $\varepsilon > 0$ is given a priori. If E^* and G^* are the images of the sets E and G under the transformation, then

$$m^*E^* \leqslant mG^* = |\Delta| \cdot mG < |\Delta| \cdot (m^*E + \varepsilon),$$

from which, since ε is arbitrary, it follows that

$$m^*E^* \leqslant |\Delta| \cdot m^*E.$$

Since an arbitrary non-degenerate transformation can be replaced by a sequence of simpler ones, where the determinant of the resultant transformation equals the product of the determinants of the component transformations, we have that

$$m^*E \leqslant |\Delta| \cdot m^*E$$

for every non-degenerate transformation.

But the transformation inverse to a non-degenerate affine transformation is also a non-degenerate affine transformation with determinant $\frac{1}{\Delta}$, where in the transition to this transformation, the roles of the sets E and E^* are interchanged. This means, according to what has been proved, that

$$m^*E \leqslant \frac{1}{|\Delta|} \cdot m^*E^*,$$

from which the theorem follows.

Further, we note the perfectly obvious fact that as the result of an affine transformation, a closed set goes over again into a closed set. From this we obtain the following theorem without difficulty.

THEOREM 8. *Under a non-degenerate affine transformation, the inner measure of the image E^* of a bounded set E satisfies the relation*

$$m_*E^* = |\Delta| \cdot m_*E.$$

Returning now to the case of an isometry, we obtain the following fundamental result.

THEOREM 9. *As the result of an isometry, neither the outer nor the inner measures of a bounded set change. In particular, a measurable set A goes over into a measurable set and, furthermore, the latter set has the same measure as A.*

Although the discussion was carried out for the two-dimensional case, we hope the reader will not find it difficult to make the transition to spaces of a greater number of dimensions.

§5. THE RELATION BETWEEN THE MEASURE OF A PLANE SET AND THE MEASURES OF ITS SECTIONS

DEFINITION. Let E be a plane set consisting of points (x, y). The one-dimensional set $E(x_0)$, consisting of numbers y such that $(x_0, y) \in E$, is called the *section* of the set E by the straight line $x = x_0$.

It is easy to see that the relations

$$A \subset B, \quad S = \sum_{k=1}^{\infty} E_k, \quad P = \prod_{k=1}^{\infty} E_k, \quad R = A - B$$

imply the relations

$$A(x) \subset B(x), \quad S(x) = \sum_{k=1}^{\infty} E_k(x), \quad P(x) = \prod_{k=1}^{\infty} E_k(x),$$

$$R(x) = A(x) - B(x).$$

Further, if F is a closed set then $F(x)$ is also closed, and if G is open then $G(x)$ is also open.[9] Finally, every section of a bounded set is a bounded (or void) set.

THEOREM 1. *Let E be a measurable plane set, contained in an open rectangle $R(a<x<b, c<y<d)$. Then*

1. *The set $E(x)$ is measurable for almost all $x \in (a, b)$;*

2. *If Δ is the set of those $x \in (a, b)$ for which $E(x)$ is measurable ($m\Delta = b - a$), then the function $mE(x)$ is measurable on the set Δ;*

3. *The formula*

$$mE = \int_{\Delta} mE(x)\,dx$$

is valid.

We shall prove this theorem, considering successively more and more complicated sets.

1. Let E be the *closed rectangle* $Q(\alpha \leqslant x \leqslant \beta, \gamma \leqslant y \leqslant \delta)$, where $a<\alpha \leqslant \beta<b$, $c<\gamma \leqslant \delta<d$. If $x \in (a, \alpha)$ or $x \in (\beta, b)$, then the set $Q(x)$ is void and hence measurable, with $mQ(x) = 0$. If, however, $\alpha \leqslant x \leqslant \beta$, then $Q(x) = [\gamma, \delta]$, so that in this case also $Q(x)$ is measurable and $mQ(x) = \delta - \gamma$. It is clear from this that the set Δ referred to in the theorem is simply the entire interval (a, b), where the function $mQ(x)$ turns out to be a step function on this interval and, therefore, measurable. Finally,

$$\int_a^b mQ(x)\,dx = \int_a^{\beta} mQ(x)\,dx = \int_a^{\beta} (\delta - \gamma)\,dx = (\beta - \alpha)(\delta - \gamma) = mQ,$$

which proves the theorem for this simple case.

2. Now let $E = G$, where G is an *open* set. Then $G(x)$ is also an open set and therefore $G(x)$ is measurable for every $x \in (a, b)$ [i.e. $\Delta = (a, b)$]. Further, the set G is representable in the form

$$G = \sum_{k=1}^{\infty} Q_k,$$

where the $Q_k = Q_k(\alpha_k \leqslant x \leqslant \beta_k, \ \gamma_k \leqslant y \leqslant \delta_k)$ are closed squares without common interior points. Then

$$G(x) = \sum_{k=1}^{\infty} Q_k(x).$$

If x is different from all the points α_k and β_k, then

$$mG(x) = \sum_{k=1}^{\infty} mQ_k(x). \qquad (*)$$

In fact, the inequality

$$mG(x) \leqslant \sum_{k=1}^{\infty} mQ_k(x)$$

---

[9] $G(x)$ must be considered to be a *linear* set.

is obvious. On the other hand, $Q_k(x)$ is either the void set or a segment, where the segments $Q_k(x)$ are pairwise without common interior points. We denote the set of all interior points of the set $Q_k(x)$ by $U_k(x)$. If $Q_k(x)$ is void, then the same is true of $U_k(x)$; but if $Q_k(x)$ is a segment, then $U_k(x)$ is the interval obtained from $Q_k(x)$ by removing its endpoints. This means we always have $mU_k(x) = mQ_k(x)$. But $G(x) \supset \sum_{k=1}^{\infty} U_k(x)$. Therefore,

$$mG(x) \geqslant m\left[\sum_{k=1}^{\infty} U_k(x)\right] = \sum_{k=1}^{\infty} mU_k(x),$$

from which (*) follows.

If a denumerable set S of points $\{\alpha_k\} + \{\beta_k\}$ is removed from (a, b), then $mG(x)$ will be the sum of a series of measurable functions on the remaining set. This implies that $mG(x)$ is measurable on $(a, b) - S$ and hence also on (a, b). But a series of positive measurable functions can be integrated term by term. Since the integrals over the interval (a, b) coincide with the corresponding integrals over the set $(a, b) - S$, we have that

$$\int_a^b mG(x)\,dx = \sum_{k=1}^{\infty} \int_a^b mQ_k(x)\,dx = \sum_{k=1}^{\infty} mQ_k = mG,$$

which proves the theorem for an open set.

3. Let $E = F$, where F is a closed set. Then $F = R - G$, where G is the open complement of F relative to the rectangle R. In this case

$$F(x) = R(x) - G(x),$$

where all sets written down here are measurable. But then

$$mF(x) = (d - c) - mG(x),$$

and the function $mF(x)$ is measurable on (a, b). Integrating the last equality, we find that

$$\int_a^b mF(x)\,dx = (b - a)(d - c) - \int_a^b mG(x)\,dx = mR - mG = mF,$$

and the theorem is proved for a closed set.

4. Let E be a set of type F_σ. Then $E(x)$ is also a set of type F_σ so that $E(x)$ is measurable for all $x \in (a, b)$. Let

$$E = \Phi_1 + \Phi_2 + \Phi_3 + \ldots,$$

where the Φ_i are closed sets. Setting

$$\Phi_1 + \Phi_2 + \ldots + \Phi_n = F_n,$$

we can write E in the form

$$E = F_1 + F_2 + F_3 + \ldots,$$

where $F_1 \subset F_2 \subset F_3 \ldots$. In this case,

$$mE = \lim_{n \to \infty} mF_n.$$

On the same basis,

$$mE(x) = \lim_{n \to \infty} mF_n(x),$$

from which the measurability of the function $mE(x)$ follows. Since all the functions $mF_n(x)$ are bounded by one number [$F_n(x)$ lies in (c, d) and therefore $mF_n(x) \leqslant d - c$], we are permitted to pass to the limit under the integral sign, i.e.

$$\int_a^b mE(x)\,dx = \lim_{n\to\infty} \int_a^b mF_n(x)\,dx = \lim_{n\to\infty} mF_n = mE.$$

This proves the theorem for F_σ sets.

5. Let E be a set of type G_δ. Then $B = R - E$ is a set of type F_σ.[10] The measurability of $E(x)$ for every $x \in (a, b)$ is trivial because $E(x)$ is a G_δ. Furthermore,

$$mE(x) = d - c - mB(x),$$

from which it follows that $mE(x)$ is a measurable function and that

$$\int_a^b mE(x)\,dx = (b-a)\,(d-c) - \int_a^b mB(x)\,dx = mR - mB = mE,$$

which again proves all assertions of the theorem.

We note that $\Delta = (a, b)$ in all cases considered.

6. Finally, let E be an arbitrary measurable set. We shall construct a set A of type F_σ and a set B of type G_δ for which

$$A \subset E \subset B,\ mA = mE = mB.$$

In this connection, we may assume that $B \subset R$. According to what has been proved above, the functions $mA(x)$ and $mB(x)$ are measurable on (a, b) and

$$mE = \int_a^b mA(x)\,dx = \int_a^b mB(x)\,dx.$$

But $A(x) \subset E(x) \subset B(x)$. This means that $mA(x) \leqslant mB(x)$ and then it follows from the equality of the integrals of these functions that they coincide almost everywhere.

Let $\mathscr{E} = (a, b)$ and

$$\Delta_0 = \underset{x}{\mathscr{E}}\,[mA(x) = mB(x)] \qquad (m\Delta_0 = b - a).$$

Since $mA(x) \leqslant m_*E(x) \leqslant m^*E(x) \leqslant mB(x)$, the set $E(x)$ is measurable for $x \in \Delta_0$. Further, the function $mA(x)$, being measurable on (a, b), is also measurable on Δ_0, and on Δ_0 the measure $mE(x)$ coincides with $mA(x)$, so that $mE(x)$ is a function measurable on Δ_0. Now if we introduce the set Δ consisting of all $x \in \mathscr{E}$ for which $E(x)$ is measurable, it turns out that $\Delta_0 \subset \Delta \subset \mathscr{E}$, from which it follows that $m\Delta = b - a$ and the function $mE(x)$ (being measurable on Δ_0) is measurable on Δ. Finally,

$$\int_\Delta mE(x)\,dx = \int_{\Delta_0} mA(x)\,dx = \int_a^b mA(x)\,dx = mA = mE.$$

---

[10] In fact, if $E = \prod_{n=1}^{\infty} G_n$, then $R - E = \sum_{n=1}^{\infty} (R - G_n)$. But $R - G_n = R \cdot CG_n$. The set CG_n is closed and R is an open rectangle and, therefore, R is a set of type F_σ because every open set which is the sum of closed squares is of type F_σ. This means that $R - G_n$ is an F_σ and then $R - E$ is an F_σ also.

This completes the proof of the theorem.

COROLLARY 1. *If E is a plane set of measure zero, then almost all its sections are also sets of measure zero.*

COROLLARY 2. *If almost all sections of a plane measurable set E are sets of measure zero, then also $mE = 0$.*

Theorem 1 carries over to the multi-dimensional case, as do all the other results of this chapter. We give its formulation for the three-dimensional case.

DEFINITION. If E is a set of points (x, y, z) in three-dimensional space, then its *section* by the plane $x = x_0$ is the set of those points (y, z) such that $(x_0, y, z) \in E$.

THEOREM 2. *Let E be a measurable set in three-dimensional space, contained in the parallelepiped $R(a < x < b, c < y < d, e < z < f)$. Then*

1) *the plane set $E(x)$ is measurable for almost all $x \in (a, b)$;*

2) *if Δ is the set of those $x \in (a, b)$ for which $E(x)$ is measurable ($m\Delta = b - a$), then the function $mE(x)$ is measurable on Δ;*

3) *the formula*

$$mE = \int_{\Delta} mE(x)\,dx$$

is valid.

CHAPTER XII

MEASURABLE FUNCTIONS OF SEVERAL VARIABLES AND THEIR INTEGRATION

§1. MEASURABLE FUNCTIONS. EXTENSION OF CONTINUOUS FUNCTIONS

Having generalized the concept of measurable set to the multi-dimensional case, we can now construct the theory of measurable functions of several variables without any difficulty. We set down the following, known, definition at the basis of the theory.

DEFINITION. The function $f(M)$, where $M(x_1, x_2, \ldots, x_n)$ is a point of the n-dimensional space R_n, defined on some set E, is said to be *measurable* on the set E if the set E itself and the sets $E(f > a)$ are measurable for arbitrary real a.

Once we have this definition at our disposal, we can extend the content of §§1, 2, 3, Chapt. IV, to the multi-dimensional case with no changes whatsoever—except that Theorem 8, §1 assumes the following form.

THEOREM. *A function $f(M)$ defined and continuous on a closed parallelepiped is measurable.*

In general, in the transition from the one-dimensional to the multi-dimensional case, the role of segments will be taken over by closed parallelepipeds. The proof of the theorem formulated above does not differ from that of the theorem introduced in Chapter IV. Note, by the way, that the same reasoning as that introduced in Chapt. IV to prove Theorem 8, §1, establishes the validity of the following lemma.

LEMMA 1. *If $f(M)$ is a function defined and continuous on the closed set F, then the sets $F(f \geqslant a)$ and $F(f \leqslant a)$ are closed for arbitrary real a.*

It is more complicated to extend the content of §4, Chapt. IV, to the multi-dimensional case. The proof of the Borel and Luzin theorems, given in that section, were based on Lemma 2 of the same section, the proof of which used, in an essential way, the fact that we were dealing with the one-dimensional case. Therefore, we shall now have to modify our discussion somewhat.

LEMMA 2. *Let E be any non-void set, contained in the n-dimensional space R_n. If we denote by $r(M)$ the distance from an arbitrary point $M \in R_n$ to the set E, $r(M) = \rho(M, E)$, then the function $r(M)$ will be continuous in the entire space.*

To prove this, we choose two points M and N in R_n. By the very definition of the concept of distance from a point to a set, having taken an arbitrary $\varepsilon > 0$, we can find a point A in the set E such that $\rho(M, A) < r(M) + \varepsilon$. But then

$$r(N) \leqslant \rho(N,A) \leqslant \rho(N,M) \leqslant \rho(M,A) < \rho(N,M) + r(M) + \varepsilon.$$

Since ε is arbitrary, it follows that

$$r(N) - r(M) \leqslant \rho(N,M),$$

and since the points M and N are interchangeable, we also have

$$|r(N)-r(M)| \leqslant \rho(N,M),$$

which proves the theorem.

LEMMA 3. *Assume the closed sets $F_1, F_2, \ldots, F_m$ are pairwise disjoint. If the function $\varphi(M)$, defined on the set*

$$F = F_1 + F_2 + \ldots + F_m,$$

is constant on each of the sets F_k, then there exists a function $\psi(M)$, defined and continuous in the entire space R_n which coincides with $\varphi(M)$ at points of the set F and is such that for all $M \in R_n$,

$$\min_F \{\varphi(M)\} \leqslant \psi(M) \leqslant \max_F \{\varphi(M)\}.$$

Proof. Let $r_k(M)$ denote the distance from the point $M \in R_n$ to the set $F_k (k = 1, 2, \ldots, m)$. In view of the fact that the set F_k is closed, the function $r_k(M)$ vanishes *only* for points in the set F_k.

Having noted this, we denote by a_k that value which the function $\varphi(M)$ assumes on the set F_k, and introduce the function $\psi(M)$, which is equal to $\varphi(M)$ for $M \in F$ and to

$$\psi(M) = \frac{\dfrac{a_1}{r_1(M)} + \dfrac{a_2}{r_2(M)} + \cdots + \dfrac{a_m}{r_m(M)}}{\dfrac{1}{r_1(M)} + \dfrac{1}{r_2(M)} + \cdots + \dfrac{1}{r_m(M)}}$$

for $M \bar{\in} F$. Using the preceding lemma, the reader can easily verify that the function $\psi(M)$ is the one required.

Now we shall prove the theorem which will take the place of Lemma 2, §4, Chapt. IV, and which is of great interest in itself.

THEOREM. *Let Φ be a closed set lying in R_n and let $\varphi(M)$ be a function, defined, bounded and continuous*[1] *on the set Φ. Then there exists a function $\psi(M)$, defined and continuous on the entire space R_n which coincides with $\varphi(M)$ at points of the set Φ and is such that*

$$|\psi(M)| \leqslant \sup_{\Phi} \{|\varphi(M)|\}$$

for all $M \in R_n$.

Proof.[2] Set $\varphi_0(M) = \varphi(M)$ and let

$$A_0 = \sup_{\Phi} \{|\varphi_0(M)|\}.$$

Introduce the sets

$$F_- = \Phi\left(\varphi \leqslant -\frac{A_0}{3}\right), \quad F_+ = \Phi\left(\varphi_0 \geqslant \frac{A_0}{3}\right).$$

By Lemma 1, both these sets are closed (where one of them can be void). By Lemma 2, we can construct a function $\psi_0(M)$, defined and continuous on the entire space R_n, equal to $-\frac{A_0}{3}$ on the set F_-, equal to $\frac{A_0}{3}$ on the set F_+, and such that

---

[1] For greater generality, we do not assume that the set Φ is bounded. Otherwise, as is known from elementary analysis, the boundedness of the function $\varphi(M)$ would follow from its continuity and $\varphi(M)$ would attain its maximum and minimum values.

[2] This proof is adopted from P. S. Aleksandrov's book *Introduction to the General Theory of Sets and Functions*, Gostekhizdat, 1948; *Theory of Functions*, Chelsea (in preparation).

$$-\frac{A_0}{3} \leqslant \psi_0(M) \leqslant \frac{A_0}{3}$$

for all $M \in R_n$.

Consider the difference

$$\varphi_1(M) = \varphi_0(M) - \psi_0(M),$$

which is defined only on the set Φ and is a bounded, continuous function on this set. It is not difficult to see that[3]

$$|\varphi_1(M)| \leqslant \tfrac{2}{3} A_0.$$

We set

$$A_1 = \sup_{\Phi} \{|\varphi_1(M)|\}$$

and construct a function $\psi_1(M)$ which plays the same role with respect to $\varphi_1(M)$ that $\psi_0(M)$ plays with respect to $\varphi_0(M)$. In other words, the function $\psi_1(M)$ will possess the following properties: It is defined on the entire space; it is continuous everywhere; it satisfies the inequality

$$-\frac{A_1}{3} \leqslant \psi_1(M) \leqslant \frac{A_1}{3}$$

for all M; and it is such that

$$|\varphi_1(M) - \psi_1(M)| \leqslant \tfrac{2}{3} A_1$$

for $M \in \Phi$. We emphasize that $A_1 \leqslant \frac{2}{3} A_0$. Having constructed such a $\psi_1(M)$, we set

$$\varphi_2(M) = \varphi_1(M) - \psi_1(M) \qquad (M \in \Phi)$$

and continue this process. As a result, we construct two sequences of continuous functions $\{\varphi_k(M)\}$ and $\{\psi_k(M)\}$, where the $\varphi_k(M)$ are defined only on Φ and the $\psi_k(M)$ are defined on all of R_n. These sequences will possess the following properties:

$$\varphi_k(M) = \varphi_{k-1}(M) - \psi_{k-1}(M) \qquad (M \in \Phi, k \geqslant 1)$$

$$-\frac{A_k}{3} \leqslant \psi_k(M) \leqslant \frac{A_k}{3} \qquad (M \in R_n)$$

$$|\varphi_k(M) - \psi_k(M)| \leqslant \tfrac{2}{3} A_k \qquad (M \in \Phi).$$

In this connection, we denote the quantity

$$\sup_{\Phi} \{|\varphi_k(M)|\}$$

by A_k, so that $A_k \leqslant \frac{2}{3} A_{k-1}$, from which it follows that $A_k \leqslant (\frac{2}{3})^k A_0$.

Having done everything stated above, we now form the infinite series

$$\psi_0(M) + \psi_1(M) + \psi_2(M) + \ldots . \qquad (*)$$

The terms of this series are continuous. Furthermore, by the inequality $|\psi_k(M)| \leqslant \left(\frac{2}{3}\right)^k \frac{A_0}{3}$, this series converges uniformly. Therefore, its sum $\psi(M)$ is a continuous function. This is the required function. In fact, it admits of the estimate

---

[3] In fact, if $M \in F_+$, then $\psi_0(M) = \frac{A_0}{3} \leqslant \varphi_0(M) \leqslant A_0$. The same is true for $M \in F_-$. If, however, $M \in \Phi - (F_- + F_+)$, then $-\frac{A_0}{3} \leqslant \psi_0(M) \leqslant \frac{A_0}{3}$ and $-\frac{A_0}{3} \leqslant \varphi_0(M) \leqslant \frac{A_0}{3}$.

$$|\psi(M)| \leqslant \sum_{k=0}^{\infty} \left(\frac{2}{3}\right)^k \frac{A_0}{3} = A_0.$$

Therefore, it only remains to verify that for $M \in \Phi$, $\psi(M) = \varphi(M)$. For this purpose, we note that for $M \in \Phi$,

$$\psi_k(M) = \varphi_k(M) - \varphi_{k+1}(M).$$

Hence, the partial sum $S_p(M) = \psi_0(M) + \psi_1(M) + \ldots + \psi_{p-1}(M)$ of the series (*), for $M \in \Phi$, can be written as

$$S_p(M) = [\varphi_0(M) - \varphi_1(M)] + [\varphi_1(M) - \varphi_2(M)] + \ldots + \\ + [\varphi_{p-1}(M) - \varphi_p(M)] = \varphi(M) - \varphi_p(M).$$

But since

$$|\varphi_p(M)| \leqslant (\tfrac{2}{3})^p A_0,$$

we have that

$$\lim_{n \to \infty} S_p(M) = \varphi(M) \qquad (M \in \Phi)$$

and this means that $\psi(M) = \varphi(M)$ for $M \in \Phi$. This completes the proof of the theorem.

Using this theorem, we can assume, without further explanation, that the entire content of §4, Chapt. IV, has been extended to the multi-dimensional case.[4]

§2. THE LEBESGUE INTEGRAL AND ITS GEOMETRIC INTERPRETATION

In the definition of the Lebesgue integral in Chapters V and VI, we did not depend anywhere on the fact that we were dealing with functions of one variable, except in §§4 and 5 of Chapt. V. Therefore, the entire content of these chapters extends to the multi-dimensional case. We shall not go into detail here. We only note that to denote the integral over a set E, lying in n-dimensional space, we will use one of the designations

$$\int_E f(M)\,dw, \qquad \underbrace{\int\int \ldots \int}_{E} f(x_1, x_2, \ldots x_n)\,dx_1\,dx_2 \ldots dx_n.$$

Let us stop to consider the geometric meaning of the integral.

DEFINITION. Let $f(M) = f(x_1, x_2, \ldots, x_n)$ be a non-negative function, defined on the set E. We shall call the set $S = S(E,f)$ of those points $(x_1, x_2, \ldots, x_n, z)$ of the space R_{n+1} for which

$$(x_1, x_2, \ldots, x_n) \in E, \quad 0 \leqslant z \leqslant f(x_1, x_2, \ldots, x_n)$$

the *subgraph*[5] of the function $f(M)$.

It is not difficult to see that in the case of a continuous positive function $f(x)$, defined on the closed interval $[a, b]$, the subgraph is nothing other than the *curvilinear trapezium*, bounded below by the x-axis, above by the curve $y = f(x)$, and on the sides by the straight lines $x = a$ and $x = b$. In this special case, as is known, the integral

---

[4] If Luzin's theorem were formulated in a somewhat weaker form, for instance: "*if $f(M)$ is defined, measurable and almost everywhere finite on a measurable set E, then to every $\varepsilon > 0$, there corresponds a set $E_\varepsilon \subset E$ such that $mE_\varepsilon > mE - \varepsilon$ and the function $f(M)$ is continuous on E_ε*," then we would not need the theorem of the present section for its proof; we could have used Lemma 3.

[5] This term is introduced here for the first time.

$$\int_a^b f(x)\,dx$$

represents the area of the indicated trapezium. We shall show that in the general case there is something quite analogous.

LEMMA. *If E is a measurable set in the space R_n, and the function $f(M)$ equals a positive number h at all points in E, then the subgraph $S(E,h)$ is a measurable set in the space R_{n+1} and*

$$mS(E,h) = hmE.$$

This lemma generalizes the elementary theorem that the volume of a cylinder equals the area of its base multiplied by its height. We shall prove this lemma restricting ourselves, for the sake of simplicity, to the case $n=2$.

If the set E is a rectangle $R(a \leqslant x \leqslant b, c \leqslant y \leqslant d)$, the lemma is trivial, because $S(E,h)$ is a rectangular parallelepiped $T(a \leqslant x \leqslant b, c \leqslant y \leqslant d, 0 \leqslant z \leqslant h)$. The same holds for the case where E is an *open* rectangle. If E is a bounded open set, then it can be represented in the form

$$E = \sum_{k=1}^{\infty} Q_k,$$

where the Q_k are closed squares which are pairwise without common interior points. Then

$$S(E,h) = \sum_{k=1}^{\infty} S(Q_k,h).$$

This means that

$$m^*S(E,h) \leqslant \sum_{k=1}^{\infty} mS(Q_k,h) = \sum_{k=1}^{\infty} h \cdot mQ_k = h \cdot mE.$$

On the other hand, if U_k is the set of interior points of the square Q_k,

$$S(E,h) \supset \sum_{k=1}^{\infty} S(U_k,h),$$

from which it follows that

$$m_*S(E,h) \geqslant \sum_{k=1}^{\infty} mS(U_k,h) = \sum_{k=1}^{\infty} h \cdot mU_k = h \cdot mE;$$

this proves the lemma for a bounded open set. But in this case, it is also true for a bounded closed set because a bounded closed set can be represented in the form $R - G$, where R is an open rectangle and G is a bounded open set. Finally, let E be an arbitrary measurable set. Take $\varepsilon > 0$; we can find a closed set F and a bounded open set G such that

$$F \subset E \subset G,\ mF > mE - \varepsilon,\ mG < mE + \varepsilon.$$

In this case

$$S(F,h) \subset S(E,h) \subset S(G,h),$$

from which it follows, according to what has already been proved, that

$$h \cdot mF \leqslant m_*S(E,h) \leqslant m^*S(E,h) \leqslant h \cdot mG$$

and, a fortiori, that

$$h(mE - \varepsilon) \leqslant m_*S(E,h) \leqslant m^*S(E,h) \leqslant h(mE + \varepsilon).$$

Since ε is arbitrary, this proves the lemma.

THEOREM. *If $f(M)$ is a bounded, measurable, non-negative function, defined on the set E, then its subgraph $S(E,f)$ is measurable and*

$$mS(E,f) = \int_E f(M)\,dw.$$

In fact, let $0 \leqslant f(M) < A$. Subdivide the interval $[0, A]$ by means of the points $z_0 = 0 < z_1 < z_2 < \ldots < z_n = A$ and form the Lebesgue sets

$$e_i = E(z_i \leqslant f < z_{i+1}).$$

It is not difficult to see that

$$S(E,f) = \sum_{i=0}^{n-1} S(e_i,f),$$

where the sets $S(e_i,f)$ are pairwise disjoint. It follows that

$$\sum_{i=0}^{n-1} m_* S(e_i,f) \leqslant m_* S(E,f) \leqslant m^* S(E,f) \leqslant \sum_{i=0}^{n-1} m^* S(e_i,f).$$

On the other hand,

$$S(e_i, z_i) \subset S(e_i,f) \subset S(e_i, z_{i+1}),$$

from which, applying the preceding theorem, we find that

$$z_i m e_i \leqslant m_* S(e_i,f), \quad m^* S(e_i,f) \leqslant z_{i+1} m e_i.$$

Thus,

$$\sum_{i=0}^{n-1} z_i m e_i \leqslant m_* S(E,f) \leqslant m^* S(E,f) \leqslant \sum_{i=0}^{n-1} z_{i+1} m e_i.$$

We complete the proof by refining the subdivision and taking the limit.

§3. FUBINI'S THEOREM

We shall now consider the problem of reducing a multiple integral to a single integral.

LEMMA. *If $f(x,y)$ is a measurable function, defined in the rectangle $R(a \leqslant x \leqslant b, c \leqslant y \leqslant d)$, then this function considered as a function of y alone is measurable on $[c,d]$ for all $x \in [a,b]$.*

Proof. Suppose first that $f(x,y)$ is bounded in R. There exists a sequence of continuous functions $g_n(x,y)$ such that

$$\lim_{n\to\infty} g_n(x,y) = f(x,y) \tag{1}$$

almost everywhere in R. We denote the set of those points $(x,y) \in R$ for which (1) is not satisfied by E. Since $mE = 0$, it follows [see Corollary 1 in §5, Chapt. XI] that

$$mE(x) = 0 \tag{2}$$

for all x in $[a,b]$, where $E(x_0)$ is the section of the set E by the straight line $x = x_0$.

Suppose x is a point in $[a,b]$ at which (2) holds. If $y \in [c,d] - E(x)$, then $(x,y) \in R - E$ and (1) is valid. In other words, for the indicated x relation (1) is satisfied almost everywhere on $[c,d]$, and then $f(x,y)$ (for the same x) is a measurable function of y.

If $f(x,y)$ is not bounded, then there exists a sequence of bounded measurable functions $f_n(x,y)$ such that

$$\lim_{n\to\infty} f_n(x,y) = f(x,y)$$

everywhere on R. By what we have proved, for every n the set e_n of those $x \in [a,b]$, for which $f_n(x,y)$ is not a measurable function on $[c,d]$, has measure zero. Suppose $e = e_1 + e_2 + e_3 + \dots$. If $x \in [a,b] - e$, then all the functions $f_n(x,y)$ are measurable on $[c,d]$. But then $f(x,y)$ is also measurable.

THEOREM 1 (G. FUBINI). *Let $f(x,y)$ be a summable function defined on the rectangle $R(a \leqslant x \leqslant b, c \leqslant y \leqslant d)$.*

Then:

1) *the function $f(x,y)$ considered as a function of y alone will be summable on $[c,d]$ for almost all $x \in [a,b]$;*

2) *if Δ is the set of those $x \in [a,b]$ for which $f(x,y)$ is summable on $[c,d]$ ($m\Delta = b-a$), then the function*

$$\int_c^d f(x,y)\,dy$$

is summable on Δ;

3) *the formula*

$$\iint_R f(x,y)\,dx\,dy = \int_\Delta dx \int_c^d f(x,y)\,dy$$

is valid.

Proof. We first consider the case when $f(x,y)$ is a bounded non-negative function,

$$0 \leqslant f(x,y) \leqslant H.$$

Its subgraph $A = S(R,f)$ will be a measurable set and

$$mA = \iint_R f(x,y)\,dx\,dy.$$

We denote the set of all those $x \in [a,b]$ for which $f(x,y)$ is a measurable function of y on $[c,d]$ by Δ'. According to the lemma, $m\Delta' = b-a$. Since, for arbitrary $x_0 \in [a,b]$, the section $A(x_0)$ of the set A by the plane $x = x_0$ is simply the subgraph of the function $f(x_0,y)$, we have (by the theorem in §2) that for $x \in \Delta'$ the set $A(x)$ is measurable and

$$mA(x) = \int_c^d f(x,y)\,dy. \tag{3}$$

On the other hand, in virtue of the results of §5, Chapt. XI, the function $mA(x)$ is measurable on the set Δ'' of those $x \in [a,b]$ for which $A(x)$ is measurable and

$$mA = \int_{\Delta''} mA(x)\,dx.$$

Since $\Delta' \subset \Delta''$[6] and $m\Delta' = m\Delta''$, $mA(x)$ is measurable on Δ' also. Hence integral (3), which coincides with $mA(x)$ on Δ', will be a measurable and (in virtue of the fact that it is bounded) summable function of x on Δ'. Furthermore,

---

[6] It can be proved that actually $\Delta' = \Delta''$.

$$mA = \int_{\Delta'} mA(x)\,dx = \int_{\Delta'} dx \int_c^d f(x,y)\,dy.$$

It remains to note that a bounded function $f(x,y)$, considered as a function of y alone will be summable if and only if it is measurable. Therefore, the set Δ' which we introduced above coincides with the set Δ referred to in the formulation of the theorem. We have thus proved the theorem for the case of a bounded non-negative function.

Now, let us omit the assumption that $f(x,y)$ is bounded, but still assume that it is non-negative. We introduce the "truncated" function

$$f_n(x,y) = \begin{cases} f(x,y), \text{ if } f(x,y) \leqslant n \\ n, \quad \text{if } f(x,y) > n. \end{cases}$$

Then

$$\iint_R f(x,y)\,dx\,dy = \lim_{n\to\infty} \iint_R f_n(x,y)\,dx\,dy.$$

The part of the theorem already proved is applicable to the function $f_n(x,y)$. If Δ_n is the set of those x for which $f_n(x,y)$ (as a function of y) is measurable on $[c,d]$, then $m\Delta_n = b-a$, the integral $\int_c^d f_n(x,y)dy$ is measurable on Δ_n, and

$$\iint_R f_n(x,y)\,dx\,dy = \int_{\Delta_n} dx \int_c^d f_n(x,y)\,dy.$$

Let $\Delta^* = \Delta_1\Delta_2\Delta_3\ldots$. It is clear that $m\Delta^* = b-a$ and that all the functions $f_n(x,y)$ are measurable[7] for $x\in\Delta^*$, and hence $f(x,y)$ is also measurable for $x\in\Delta^*$. Furthermore,

$$\iint_R f_n(x,y)\,dx\,dy = \int_{\Delta^*} dx \int_c^d f_n(x,y)\,dy.$$

In view of the fact that

$$f_1(x,y) \leqslant f_2(x,y) \leqslant f_3(x,y) \leqslant \ldots, \lim_{n\to\infty} f_n(x,y) = f(x,y),$$

we can apply B. Levi's theorem (for fixed $x\in\Delta^*$). Therefore

$$\lim_{n\to\infty} \int_c^d f_n(x,y)\,dy = \int_c^d f(x,y)\,dy \qquad (x\in\Delta^*)$$

where, in the process, we have established the measurability of the integral

$$\int_c^d f(x,y)\,dy$$

as a function of $x\in\Delta^*$.

In view of the fact that

---

[7] As a function of y.

$$\int_c^d f_1(x,y)\,dy \leqslant \int_c^d f_2(x,y)\,dy \leqslant \int_c^d f_3(x,y)\,dy \leqslant \ldots,$$

we can apply B. Levi's theorem once more, which yields the equality

$$\lim_{n\to\infty} \int_{\Delta^*} dx \int_c^d f_n(x,y)\,dy = \int_{\Delta^*} dx \int_c^d f(x,y)\,dy.$$

We have thus found a set Δ^*, of measure $m\Delta^* = b - a$, for all points x of which the function $f(x,y)$ is measurable, where the integral $\int_c^d f(x,y)\,dy$ represents a function measurable on Δ^* and

$$\iint_R f(x,y)\,dx\,dy = \int_{\Delta^*} dx \int_c^d f(x,y)\,dy.$$

Since the integral in the left member is a finite quantity, the integral $\int_c^d f(x,y)\,dy$ represents a function which is not only measurable but also summable on the set Δ^*. And since a summable function is finite almost everywhere, the integral $\int_c^d f(x,y)\,dy$ is finite for almost all x in Δ^*. This shows that the function $f(x,y)$, as a function of y only, is summable on $[c,d]$ for almost all x in Δ^*. We denote by Δ^{**} the set of those x in Δ^* for which $f(x,y)$ is summable on $[c,d]$. Since $m\Delta^{**} = m\Delta^*$,

$$\iint_R f(x,y)\,dx\,dy = \int_{\Delta^{**}} dx \int_c^d f(x,y)\,dy.$$

Having noted this, we denote by Δ the set of *all* x in $[a,b]$ for which $f(x,y)$ is a summable function of the argument y. It is clear that $\Delta^{**} \subset \Delta \subset [a,b]$; therefore, the integral

$$\int_c^d f(x,y)\,dy$$

is also summable on the set Δ because $m(\Delta - \Delta^{**}) = 0$. Then

$$\int_{\Delta^{**}} dx \int_c^d f(x,y)\,dy = \int_{\Delta} dx \int_c^d f(x,y)\,dy,$$

and the theorem is proved for the case when $f(x,y)$ is non-negative, but not necessarily bounded.

Finally, let us go over to the general case, assuming that $f(x,y)$ is an arbitrary summable function. As usual, we set

$$f_+(x,y)=\begin{cases} f(x,y), & \text{if } f(x,y)\geqslant 0;\\ 0 & \text{if } f(x,y)<0;\end{cases}$$
$$f_-(x,y)=\begin{cases} 0 & \text{if } f(x,y)\geqslant 0;\\ -f(x,y), & \text{if } f(x,y)<0.\end{cases}$$

These functions are summable and non-negative; therefore, we can apply the part of the theorem already proved to them. We denote by Δ_+ and Δ_- the sets of those x in $[a,b]$ for which the functions $f_+(x,y)$ and $f_-(x,y)$, respectively, are summable on $[c,d]$. According to what has already been proved, $m\Delta_+ = m\Delta_- = b-a$, where the integrals

$$\int_c^d f_+(x,y)\,dx, \qquad \int_c^d f_-(x,y)\,dy$$

are summable on Δ_+ and Δ_-, respectively, and

$$\iint_R f_+(x,y)\,dx\,dy = \int_{\Delta_+} dx \int_c^d f_+(x,y)\,dy,$$

$$\iint_R f_-(x,y)\,dx\,dy = \int_{\Delta_-} dx \int_c^d f_-(x,y)\,dy.$$

Set $\Delta = \Delta_+ \Delta_-$. Then $m\Delta = b-a$, and the integrals in the right members of the preceding equalities can be extended not only to Δ_+ and Δ_- but also to Δ. Then the memberwise subtraction of these equations yields

$$\iint_R f(x,y)\,dx\,dy = \int_{\Delta} dx \int_c^d f(x,y)\,dy.$$

It remains to note that the set $\Delta = \Delta_+ \Delta_-$ consists *only* of such $x \in [a,b]$ for which $f(x,y)$, as a function of y, is summable on $[c,d]$ and that *all* such x appear in Δ.

This completes the proof of the Fubini theorem.

Remarks. 1. If the function $f(x,y)$ is defined on an arbitrary measurable set E, rather than on a rectangle, then we must enclose this set in the rectangle R and extend the definition of $f(x,y)$ by setting it equal to zero on the set $R-E$. This reduces the problem to the case already considered.

2. It is not difficult to see that if $f(x,y)$ is a measurable,[8] non-negative function, defined on the rectangle $R(a\leqslant x\leqslant b, c\leqslant y\leqslant d)$, then integral (3), considered on the set Δ[9] of those $x\in[a,b]$ for which $f(x,y)$ is a measurable function of y ($m\Delta = b-a$), will be a measurable function of x.

In fact, if $f(x,y)$ is bounded, then it is summable and both these assertions are already contained in Theorem 1. If, however, $f(x,y)$ is not bounded, one must simply take the limit of the truncated functions and use the theorem of B. Levi.

---

[8]It is not assumed here that $f(x,y)$ is summable.

[9]Here, the integral $\int_c^d f(x,y)\,dy$ can equal $+\infty$ on a set of positive measure.

Theorem 1 admits of the following generalization which can be proved by following the same line of reasoning used in proving the theorem itself.

THEOREM 2. *Let* $f(x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_m)$ *be a summable function, defined on the* $(n+m)$*-dimensional parallelepiped*

$$R_{n+m}(a_1 \leqslant x_1 \leqslant b_1, \ldots, a_n \leqslant x_n \leqslant b_n; c_1 \leqslant y_1 \leqslant d_1, \ldots, c_m \leqslant y_m \leqslant d_m).$$

Let $R'_n = R'_n(a_1 \leqslant x_1 \leqslant b_1, \ldots, a_n \leqslant x_n \leqslant b_n)$ *and* $R''_m = R''_m(c_1 \leqslant y_1 \leqslant d_1, \ldots, c_m \leqslant y_m \leqslant d_m)$. *Then:*

1. *The function* $f(x_1, \ldots, x_n; y_1, \ldots, y_m)$ *is summable on* R''_m *for almost all points* $(x_1, x_2, \ldots, x_n) \in R'_n$.

2. *If* Δ_n *is the set of these points* $(m\Delta_n = mR'_n)$, *then the integral*

$$\underbrace{\int\int \cdots \int}_{R''_m} f(x_1, \ldots, x_n; y_1, \ldots, y_m)\, dy_1 \ldots dy_m$$

is summable on Δ_n.

3. *The following formula is valid*:

$$\underbrace{\int\int \cdots \int}_{R_{n+m}} f(x_1, \ldots, x_n; y_1, \ldots, y_m)\, dx_1 \ldots dx_n\, dy_1 \ldots dy_m =$$

$$= \underbrace{\int \cdots \int}_{\Delta_n} dx_1 \ldots dx_n \underbrace{\int \cdots \int}_{R''_m} f(x_1, \ldots, x_n; y_1, \ldots, y_m)\, dy_1 \ldots dy_m.$$

§4. INTERCHANGING THE ORDER OF INTEGRATION

We introduce the following generalization of the integral symbol. Let $f(M)$ be a function, defined and summable on a measurable set A. If E is an arbitrary measurable set containing A, $E \supset A$, and such that

$$m(E - A) = 0,$$

then we shall set, by definition,

$$\int_E f(M)\, dw = \int_A f(M)\, dw.$$

Thus, we now allow cases where the function under the integral sign is defined only at almost all, but not necessarily at all, points in the region of integration. With this extension of the meaning of the integral symbol, the formula proved in Fubini's theorem can be written more simply as

$$\iint_R f(x, y)\, dx\, dy = \int_a^b dx \int_c^d f(x, y)\, dy.$$

Since the arguments x and y are interchangeable, we obtain the following result.

THEOREM 1. *If* $f(x, y)$ *is a function summable on the rectangle* $R(a \leqslant x \leqslant b, c \leqslant y \leqslant d)$, *then the two iterated integrals*

$$\int_a^b dx \int_c^d f(x,y)\,dy, \quad \int_c^d dy \int_a^b f(x,y)\,dx \tag{*}$$

exist[10] *and*

$$\int_a^b dx \int_c^d f(x,y)\,dy = \int_c^d dy \int_a^b f(x,y)\,dx. \tag{**}$$

It is interesting to note that the integrals (*) can exist even when $f(x,y)$ is not summable on the rectangle R.

Example. Let the function $f(x,y)$ be defined in the square $Q(-1 \leqslant x \leqslant 1, -1 \leqslant y \leqslant 1)$ by the formula

$$f(x,y) = \frac{xy}{(x^2+y^2)^2},$$

for $x^2+y^2>0$. We set $f(0,0)=0$. If we fix one of the arguments x or y, then $f(x,y)$ is a *continuous* function of the other so that

$$\int_{-1}^{+1} \frac{xy\,dy}{(x^2+y^2)^2}$$

exists, as is known, for all x in $[-1,+1]$. Being the integral of an odd function, this integral is equal to zero.

Thus

$$\int_{-1}^{+1} dx \int_{-1}^{+1} \frac{xy}{(x^2+y^2)^2} dy = \int_{-1}^{+1} dy \int_{-1}^{+1} \frac{xy}{(x^2+y^2)^2} dx = 0.$$

But at the same time, the function $f(x,y)$ is not summable on Q. In fact, if it were summable on Q, then it would also be summable on the sub-square $Q^*(0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1)$. Then, by Theorem 1, the finite integral

$$\int_0^1 dx \int_0^1 \frac{xy}{(x^2+y^2)^2}\,dy$$

must exist.

But this is not true because, for $x \neq 0$,

$$\int_0^1 \frac{xy}{(x^2+y^2)^2}\,dy = \frac{1}{2x} - \frac{x}{2(x^2+1)},$$

and this function is not summable on $(0,1]$.

In the example introduced above, despite the non-summability of the function

[10] In the extended sense just indicated: the inner integrals exist almost everywhere in the region of integration of the outer integrals.

$f(x,y)$, the formula $(*_*)$ was nevertheless valid.[11] Generally speaking, however, this may not be so.

Example. If we define $f(x,y)$ on $(0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1)$ as $f(0,0) = 0$, and

$$f(x,y) = \frac{x^2 - y^2}{(x^2 + y^2)^2}$$

for points other than the origin of coordinates, then

$$\int_0^1 dx \int_0^1 f(x,y)\,dy = \frac{\pi}{4}, \quad \int_0^1 dy \int_0^1 f(x,y)\,dx = -\frac{\pi}{4}.$$

In fact, for $x \neq 0$, we have that

$$f(x,y) = \frac{\partial}{\partial y}\left(\frac{y}{x^2 + y^2}\right)$$

and, therefore, for $x \neq 0$,

$$\int_0^1 f(x,y)\,dy = \left[\frac{y}{x^2 + y^2}\right]_0^1 = \frac{1}{x^2 + 1},$$

which yields the indicated value for the first iterated integral.

The second iterated integral is evaluated similarly.

It is important to notice that for measurable functions, *which do not change sign*, such unpleasantness cannot arise. Namely, the following theorem is valid.

THEOREM 2. *Let $f(x,y)$ be a measurable, non-negative function, defined on the rectangle $R(a \leqslant x \leqslant b,\ c \leqslant y \leqslant d)$. If one of the iterated integrals* (*) *is finite,*[12] *then $f(x,y)$ is summable on R, so that the second integral in* (*) *also exists and formula* $(*_*)$ *is valid.*

In fact, suppose the first of the integrals (*) is finite and that the double integral of the function $f(x,y)$ over the rectangle R is equal to $+\infty$. Set

$$f_n(x,y) = \begin{cases} f(x,y), & \text{if } f(x,y) \leqslant n \\ n, & \text{if } f(x,y) > n. \end{cases}$$

Then

$$\lim_{n \to \infty} \iint_R f_n(x,y)\,dx\,dy = +\infty,$$

and, for sufficiently large n,

---

[11] G. M. Fichtenholz, *Sur une fonction de deux variables sans intégrale double*, Fund. Math., 6, 1924, pp. 30-36, constructed an example of a function $f(x,y)$, non-summable on $R(a \leqslant x \leqslant b, c \leqslant y \leqslant d)$, such that the equality

$$\int_P dx \int_Q f(x,y)\,dy = \int_Q dy \int_P f(x,y)\,dx$$

holds for *any* measurable sets P and Q contained in $[a,b]$ and $[c,d]$, respectively.

[12] The *existence* (but not the finiteness!) of both integrals (*) follows from Remark 2 to Theorem 1 of the preceding section. Thus, formula $(*_*)$ is also true without the assumption that one of the integrals (*) is finite (but then it may take on the form $+\infty = +\infty$).

$$\iint_R f_n(x,y)\,dx\,dy > \int_a^b dx \int_c^d f(x,y)\,dy.$$

But this is an absurdity, because $f_n(x,y)$ is summable and hence

$$\iint_R f_n(x,y)\,dx\,dy = \int_a^b dx \int_c^d f_n(x,y)\,dy,$$

and $f_n(x,y) \leqslant f(x,y)$.

This completes the proof of the theorem.

COROLLARY. *If $f(x,y)$ is measurable on $R(a \leqslant x \leqslant b, c \leqslant y \leqslant d)$ and*

$$\int_a^b dx \int_c^d |f(x,y)|\,dy < +\infty,$$

then the integrals (*) *exist and formula* ($*_*$) *is valid.*

CHAPTER XIII

SET FUNCTIONS AND THEIR APPLICATIONS IN THE THEORY OF INTEGRATION

§1. ABSOLUTELY CONTINUOUS SET FUNCTIONS

In this chapter we shall concern ourselves with extending the properties of the indefinite Lebesgue integral to the multi-dimensional case. As above, we shall limit ourselves to the case of the plane, but everything discussed in the sequel can be generalized without difficulty to an arbitrary multi-dimensional space.

Let $\mathfrak{A}$ be a family of sets e, $\mathfrak{A}=\{e\}$. If to each set $e\in\mathfrak{A}$ there corresponds some number $\Phi(e)$, we say that a *set function* is defined on the family $\mathfrak{A}$. In the sequel, we shall assume that $\mathfrak{A}$ consists of all measurable sets contained in one and the same open rectangle R:

$$R(x_0<x<X, y_0<y<Y),$$

and that all set functions considered are finite.

The set function $\Phi(e)$ is said to be *additive* if for any two disjoint measurable sets e_1 and e_2, we have that

$$\Phi(e_1+e_2)=\Phi(e_1)+\Phi(e_2).$$

It is not difficult to see that such a function possesses the following property: for an arbitrary finite number of mutually disjoint measurable sets $e_1, e_2, \ldots, e_n$, the equality

$$\Phi\left(\sum_{k=1}^{n} e_k\right)=\sum_{k=1}^{n}\Phi(e_k)$$

is valid.

But if for a *countable* set of mutually disjoint measurable sets $e_1, e_2, e_3, \ldots$,[1]

$$\Phi\left(\sum_{k=1}^{\infty} e_k\right)=\sum_{k=1}^{\infty}\Phi(e_k),$$

the function $\Phi(e)$ is said to be *countably*, or *completely, additive.*

The set function $\Phi(e)$ is called *absolutely continuous* if its value tends to zero together with the measure of the set e, i.e. if to every $\varepsilon>0$ there corresponds a $\delta>0$ such that for an arbitrary measurable set $e\,(e\subset R)$ of measure $me<\delta$, we have $|\Phi(e)|<\varepsilon$.

It is clear that the value of such a function equals zero on every set of measure zero.[2]

THEOREM 1. *Every additive, absolutely continuous set function is also countably additive.*

In fact, if the sets e_1, e_2, e_3, . . . are pairwise disjoint, then

---

[1]Recall that all sets e_k are assumed to lie in R; this means their sum is bounded and hence measurable.

[2]If e_0 is the void set, then for every additive set function $\Phi(e)$, we have that $\Phi(e_0)=0$.

$$m\left(\sum_{k=n+1}^{\infty} e_k\right) \to 0 \text{ as } n \to \infty,$$

from which it also follows that

$$\Phi\left(\sum_{k=n+1}^{\infty} e_k\right) \to 0.$$

But

$$\Phi\left(\sum_{k=1}^{\infty} e_k\right) = \sum_{k=1}^{n} \Phi(e_k) + \Phi\left(\sum_{k=n+1}^{\infty} e_k\right).$$

The remainder of the proof is clear.

Let A be an arbitrary point set (contained in R). Any measurable set E satisfying the relations[3]

$$E \supset A, \quad mE = m^*A$$

is called a *measurable hull* of A.

Henceforth, unless it is explicitly stated otherwise, we shall consider only measurable hulls which are contained in R.

LEMMA 1. *If A is an arbitrary point set ($A \subset R$) and $\Phi(e)$ is an additive, absolutely continuous set function, then for any two measurable hulls E_1 and E_2 of the set A, we have*

$$\Phi(E_1) = \Phi(E_2).$$

In fact, if $E_3 = E_1E_2$, then $E_1 \supset E_3 \supset A$, from which it is clear that $mE_1 = mE_3$. This means that $m(E_1 - E_3) = 0$ and that $\Phi(E_1 - E_3) = 0$. But $E_1 = E_3 + (E_1 - E_3)$, so that $\Phi(E_1) = \Phi(E_3)$. Analogously, $\Phi(E_2) = \Phi(E_3)$ also.

Henceforth, when we are considering an additive, absolutely continuous set function $\Phi(e)$, we shall denote by $\Phi(A)$ the value of this function on the measurable hulls of the set A. Thus, the symbol $\Phi(A)$ is also meaningful when the set A is non-measurable (but, as always, $A \subset R$).

Further, we agree to use the following notation: if M is an arbitrary point in the plane and h is any positive number, then we shall denote the closed square with center at the point M, whose sides are parallel to the axes and have length equal to h, by $Q(M, h)$.

DEFINITION. Suppose $\Phi(e)$ is a set function and that M is any point of the rectangle R. If the number λ (which can equal plus or minus infinity) is such that there exists a sequence of positive numbers $\{h_n\}$ tending to zero, for which[4]

$$\lim_{n\to\infty} \frac{\Phi[Q(M, h_n)]}{h_n^2} = \lambda,$$

then we shall say that λ is a *symmetric derived number* of the function $\Phi(e)$ at the point M and write $\lambda = D_M\Phi(e)$.

It is easy to show, with the aid of the Bolzano-Weierstrass theorem, that every set function has symmetric derived numbers at every point $M \in R$.

If all symmetric derived numbers of the function $\Phi(e)$ are equal at any point M, then we shall say that the *symmetric derivative* $D_M\Phi(e)$ of $\Phi(e)$ exists at the point M_2

---

[3]It was proved in Chapter XI that every bounded measurable set has a measurable hull (of type G_δ).

[4]Recall that R is an *open* rectangle. This means that for all sufficiently small h, $Q(M, h) \subset R$, and the symbol $\Phi[Q(M, h)]$ has meaning.

and is equal to the common value of all the derived numbers. We note that the symmetric derivative could have been defined differently, namely, as the limit

$$\lim_{h\to 0} \frac{\Phi\,[Q\,(M,h)]}{h^2}.$$

However, the two definitions are equivalent.

We now prove a lemma fundamental for the entire chapter.

LEMMA 2. *Let $\Phi(e)$ be an additive, absolutely continuous set function. If at every point M of some set $A\,(A \subset R)$, there exists at least one symmetric derived number $D_M\Phi(e)$, satisfying the inequality*

$$D_M\Phi(e) \geqslant q,$$

then

$$\Phi(A) \geqslant q \cdot m^*A.$$

Proof. Let $q_0 < q$ and $\varepsilon > 0$. We find a $\delta > 0$ such that the inequality $me < 2\delta$ implies the inequality $|\Phi(e)| < \varepsilon$. In this connection, we may assume that $\delta < \varepsilon$. Further, let G be an open set such that

$$A \subset G \subset R, \quad mG < m^*A + \delta.$$

If $M \in A$, then there exists a sequence $\{h_n\}$ such that

$$h_n > 0, \; Q(M,h_n) \subset R, \; h_n \to 0$$

and

$$\lim_{n\to\infty} \frac{\Phi\,[Q\,(M,h_n)]}{h_n^2} = D_M\Phi(e) > q_0.$$

Without loss of generality, we may assume that for all n,

$$Q(M,h_n) \subset R, \; \Phi\,[Q(M,h_n)] > q_0 h_n^2.$$

It is clear that the squares $Q(M,h_n)$ cover the set A in the Vitali sense. Therefore, it is possible to select a countable set of squares $Q_1, Q_2, Q_3, \ldots$ such that

$$Q_iQ_k = 0 \quad (i \neq k), \qquad m\,[A - \sum_{k=1}^{\infty} Q_k] = 0.$$

If we set

$$S = \sum_{k=1}^{\infty} Q_k,$$

we have that

$$\Phi(S) = \sum_{k=1}^{\infty} \Phi(Q_k) > q_0 \sum_{k=1}^{\infty} mQ_k = q_0 \cdot mS.$$

By virtue of the fact that $A \subset S + (A - S)$, we have

$$m^*A \leqslant mS + m(A - S) = mS.$$

On the other hand, $S \subset G$. Therefore, $mS \leqslant mG$. This means that

$$m^*A \leqslant mS \leqslant mG < m^*A + \delta \qquad (*)$$

and

$$m(G - S) = mG - mS < \delta.$$

Furthermore, it follows from (*) that $mS = m^*A + \theta\varepsilon$, where $0 \leqslant \theta < 1$.

Having noted this, we take any measurable hull E of the set A. We may assume

that $E \subset G$. But then $E - SE \subset G - S$ and hence $m(E - SE) < \delta$, so that a fortiori $|\Phi(E - SE)| < \varepsilon$, or, what amounts to the same thing,

$$|\Phi(E) - \Phi(SE)| < \varepsilon.$$

But the inequality $m(E - SE) < \delta$ also means that

$$mSE > mE - \delta = m^*A - \delta > mG - 2\delta \geqslant mS - 2\delta.$$

Hence, $m(S - SE) < 2\delta$ and

$$|\Phi(S) - \Phi(SE)| < \varepsilon.$$

Thus,

$$|\Phi(E) - \Phi(S)| < 2\varepsilon,$$

and

$$\Phi(E) > q_0 mS - 2\varepsilon.$$

In other words, the inequality

$$\Phi(E) > q_0 [m^*A + \theta\varepsilon] - 2\varepsilon$$

is valid.

Since $\Phi(E) = \Phi(A)$, we can now complete the proof by first taking the limit as $\varepsilon \to 0$ and then passing to the limit as $q_0 \to q$.

COROLLARY 1. *If $\Phi(e)$ is an additive, absolutely continuous set function and at every point M of the set A there exists at least one symmetric derived number $D_M\Phi(e)$, satisfying the inequality*

$$D_M\Phi(e) \leqslant p,$$

then

$$\Phi(A) \leqslant pm^*A.$$

In fact, if $\Phi_1(e) = -\Phi(e)$, then at every point $M \in A$ there exists a symmetric derived number $D_M\Phi_1(e)$ such that

$$D_M\Phi_1(e) \geqslant -p.$$

Hence, by Lemma 2, $\Phi_1(A) \geqslant -pm^*A$, and this is equivalent to the inequality we wish to prove.

COROLLARY 2. *Let $\Phi(e)$ be additive and absolutely continuous. If $p < q$ and at every point M of the set A there exist two symmetric derived numbers $D'_M\Phi(e)$ and $D''_M\Phi(e)$, such that*

$$D'_M\Phi(e) < p < q < D''_M\Phi(e), \tag{*}$$

then $m^*A = 0$.

In fact, by Lemma 2 and Corollary 1, the two inequalities

$$\Phi(A) \geqslant qm^*A, \qquad \Phi(A) \leqslant pm^*A$$

are satisfied.

Therefore

$$qm^*A \leqslant pm^*A;$$

but this is possible only if $m^*A = 0$.

THEOREM 2. *If the set function $\Phi(e)$ is additive and absolutely continuous, then the finite symmetric derivative $D_M\Phi(e)$ exists at almost all points of the rectangle R.*

In fact, at every point M where the symmetric derivative does not exist, two unequal symmetric derived numbers $D'_M\Phi(e)$ and $D''_M\Phi(e)$ can be found. This means that the set A of points in R at which the symmetric derivative does not exist can be written in the form

$$A = \sum A_{p,q},$$

where $A_{p,q}$ denotes the set of points $M \in R$ at which (*) is satisfied, and the summation extends over all pairs of rational numbers (p,q), $p < q$. By Corollary 2, every term $A_{p,q}$ is a set of measure zero, and therefore $mA = 0$ also.

It remains to show that the derivative $D_M\Phi(e)$ cannot become infinite on a set of positive outer measure. But, if we denote the set of those points in R at which $D_M\Phi(e) = +\infty$ by B, then by Lemma 2,

$$\Phi(B) \geqslant qm^*B$$

for arbitrary q, from which it follows that $m^*B = 0$. We treat the set of points at which $D_M\Phi(e) = -\infty$ in a similar manner.

THEOREM 3. *Let $\Phi(e)$ be an additive, absolutely continuous set function. If R_0 is the set of those points of the rectangle R at which the symmetric derivative $D_M\Phi(e)$ exists, then this derivative is a measurable function on R_0.*

Proof. We introduce a rectangle R_1 which is wider and longer than R and defined by the inequalities

$$R_1(x_0 - 1 < x < X + 1,\ y_0 - 1 < y < Y + 1).$$

If $M \in R$ and $0 < h \leqslant 1$, then $Q(M,h) \subset R_1$. We extend the definition of the function $\Phi(e)$ by setting

$$\Phi(e) = \Phi(eR)$$

for every measurable set e contained in R_1. It is clear that the function $\Phi(e)$ remains additive and absolutely continuous even after this extension.

At every point $M \in R_0$, we have

$$D_M\Phi(e) = \lim_{n\to\infty} n^2\Phi\left[Q\left(M, \frac{1}{n}\right)\right].$$

Therefore, it is sufficient to show that for fixed $h\,(0 < h \leqslant 1)$ the function $\Phi[Q(M,h)]$ is measurable on the set R_0. Instead of this, however, we shall show that this function is *continuous* in the entire rectangle R, from which its measurability on R, and hence also on R_0, will follow.

Consider two points $M(a,b)$ and $N(a+\alpha, b+\beta)$ of the rectangle R, assuming for simplicity that $\alpha \geqslant 0$ and $\beta \geqslant 0$. It is clear that

$$Q(M,h) = Q\left(a - \frac{h}{2} \leqslant x \leqslant a + \frac{h}{2},\ b - \frac{h}{2} \leqslant y \leqslant b + \frac{h}{2}\right),$$

$$Q(N,h) = Q\left(a + \alpha - \frac{h}{2} \leqslant x \leqslant a + \alpha + \frac{h}{2},\ b + \beta - \frac{h}{2} \leqslant y \leqslant b + \beta + \frac{h}{2}\right).$$

If $\alpha < h$ and $\beta < h$, then the squares $Q(M,h)$ and $Q(N,h)$ will have intersection[5] T defined by the inequalities

$$T\left(a + \alpha - \frac{h}{2} \leqslant x \leqslant a + \frac{h}{2},\ b + \beta - \frac{h}{2} \leqslant y \leqslant b + \frac{h}{2}\right).$$

This means that

$$mT = (h - \alpha)(h - \beta)$$

and

[5]It is recommended that the reader draw a diagram.

$$m[Q(M,h)-T]=m[Q(N,h)-T]=\alpha h+\beta h-\alpha\beta<h(\alpha+\beta).$$

We take $\varepsilon>0$ and find a $\delta>0$ such that $me<\delta$ implies that $|\Phi(e)|<\varepsilon$.

If α and β are so small that $h(\alpha+\beta)<\delta$, then

$$|\Phi[Q(M,h)]-\Phi(T)|<\varepsilon,\ |\Phi[Q(N,h)]-\Phi(T)|<\varepsilon,$$

from which it follows that

$$|\Phi[Q(M,h)]-\Phi[Q(N,h)]|<2\varepsilon.$$

This completes the proof of the theorem.

The situation where the symmetric derivative $D_M\Phi(e)$ is defined only *at almost all*, and not *at all*, points of R, presents known inconveniences. In order to avoid these inconveniences, we introduce the function $\varphi(M)$ defined on all of R in the following way:

$$\varphi(M)=\begin{cases} D_M\Phi(e), & \text{if } M\in R_0 \\ 0, & \text{if } M\in R-R_0. \end{cases}$$

Here, as above, R_0 denotes the set of those points of R at which the symmetric derivative of the function $\Phi(e)$ exists.

THEOREM 4. *The function $\varphi(M)$ is summable on the rectangle R. Furthermore, for every measurable set $E\subset R$, the formula*

$$\int_E \varphi(M)\,dw=\Phi(E)$$

is valid.

Proof. Let E be a measurable set contained in R. Remove from E those points at which the symmetric derivative $D_M\Phi(e)$ does not exist, or, if it exists, it is infinite, and let A be the set of remaining points. By virtue of the fact that $m(E-A)=0$, it is clear that

$$\Phi(E)=\Phi(A).$$

Denote the set of those points M in A at which $D_M\Phi(e)\geqslant 0$ by A_+ and set $A_-=A-A_+$. Then $\Phi(A)=\Phi(A_+)+(A_-)$.

Consider the sequence of numbers

$$0,\ \frac{1}{n},\ \frac{2}{n},\ \frac{3}{n},\ \frac{4}{n},\ \ldots,\ \frac{m}{n},\ \ldots,$$

and set

$$e_k=A\left(\frac{k-1}{n}\leqslant D_M\Phi(e)<\frac{k}{n}\right)\qquad (k=1,2,3,\ldots).$$

The sets e_k are measurable, they are pairwise disjoint, and

$$\sum_{k=1}^{\infty} e_k=A_+.$$

Therefore

$$\Phi(A_+)=\sum_{k=1}^{\infty}\Phi(e_k).$$

On the other hand, the function $D_M\Phi(e)$ is measurable and non-negative on the set A_+. This means that[6]

[6]It is emphasized that at this step of the reasoning, we still cannot guarantee the finiteness of either the left or the right member of equality (*).

$$\int_{A_+} D_M\Phi(e)\,dw = \sum_{k=1}^{\infty} \int_{e_k} D_M\Phi(e)\,dw. \qquad (*)$$

By the mean value theorem,

$$\frac{k-1}{n} me_k \leqslant \int_{e_k} D_M\Phi(e)dw \leqslant \frac{k}{n}\, me_k.$$

On the other hand, by Lemma 2 and Corollary 1 to it,

$$\frac{k-1}{n} me_k \leqslant \Phi(e_k) \leqslant \frac{k}{n}\, me_k.$$

Therefore

$$\left| \Phi(e_k) - \int_{e_k} D_M\Phi(e)\,dw \right| \leqslant \frac{1}{n}\, me_k.$$

This implies the convergence of the series on the right in equation (*) and the estimate

$$\left| \Phi(A_+) - \int_{A_+} D_M\Phi(e)\,dw \right| \leqslant \frac{1}{n} \sum_{k=1}^{\infty} me_k = \frac{1}{n} mA_+.$$

Since n can be taken arbitrarily large, we have that

$$\Phi(A_+) = \int_{A_+} D_M\Phi(e)\,dw.$$

In an analogous manner we can establish a similar equation in which A_+ is simply replaced by A_-. Therefore, the function $D_M\Phi(e)$ is summable on the set A and

$$\Phi(A) = \int_A D_M\Phi(e)\,dw,$$

from which it follows that

$$\Phi(E) = \int_E D_M\Phi(e)\,dw.$$

Since, in particular, we can take $E = R$, the theorem is proved completely.

§2. THE INDEFINITE INTEGRAL AND THE DIFFERENTIATION OF IT

Let $f(M)$ be a summable function defined on an open rectangle R. We obtain a set function defined on all measurable subsets of the rectangle R by assuming that

$$\Phi(e) = \int_e f(M)\,dw$$

for every measurable set $e \subset R$. This set function is called the *indefinite integral* of the function $f(M)$. By the known properties of the Lebesgue integral, it is additive and absolutely continuous. Conversely, we showed in Theorem 4 of the preceding section that every additive, absolutely continuous set function is the indefinite integral of the function $\varphi(M)$. Therefore the following theorem is valid.

THEOREM 1. *The class of additive, absolutely continuous set functions coincides with the class of the indefinite integrals of summable functions.*

Remark. The entire content of the preceding section would be retained if the family $\mathfrak{A}$ of sets e, on which the function $\Phi(e)$ is defined, consisted of *all* measurable sets of the plane (and not only those contained in a fixed rectangle). As a result, we would have established that for an additive, absolutely continuous set function $\Phi(e)$ the symmetric derivative $D_M\Phi(e)$ exists, is measurable on every measurable set E, and is such that

$$\Phi(E) = \int_E D_M\Phi(e)\,dw$$

at almost all points of the plane.[7]

In this connection, it would be natural to define the concept of indefinite integral for every function $f(M)$, *defined on the entire plane* and summable on every measurable set. Having done this, we would have established that every absolutely continuous additive set function is the indefinite integral of its symmetric derivative. However, we could not have stated that every indefinite integral is an absolutely continuous set function, i.e. Theorem 1 would not be valid.

In fact, if (restricting ourselves to the linear case for the sake of simplicity) we set $f(x) = 0$ for $x < 1$ and $f(x) = n$ for $n \leqslant x < n+1$, then $f(x)$ will be summable on every measurable set e but its indefinite integral will not be absolutely continuous. In order to avoid such inconveniences, we required all functions considered to be defined on a fixed rectangle.

A summable function $f(M)$ defines its indefinite integral uniquely. The following proposition, which is the converse of this in a certain sense, is valid.

THEOREM 2. *If the functions $f(M)$ and $g(M)$ have the same indefinite integral, then they are equivalent.*

In fact, by hypothesis, for every measurable set e (contained in the fundamental rectangle R), we have

$$\int_e f(M)\,dw = \int_e g(M)\,dw.$$

In particular, for the set $e = E(f > g)$, we have

$$\int_e [f(M) - g(M)]\,dw = 0,$$

from which it follows (Chapt. VI, §1, Theorem 6) that

$$mE(f > g) = 0.$$

We establish, analogously, that the set $E(g < f)$ also has measure zero. This completes the proof of the theorem.

Let $f(M)$ be any summable function on R and let $\Phi(e)$ be its indefinite integral. We introduce again the function $\varphi(M)$ equal to the symmetric derivative $D_M\Phi(e)$ of the function $\Phi(e)$ where this derivative exists (i.e. almost everywhere on R) and to zero at the remaining points of R. In Theorem 4 of the preceding section we established

[7]This means that the intersection of any measurable set with the set of points at which the symmetric derivative does not exist has measure zero.

that $\Phi(e)$ is the indefinite integral of the function $\varphi(M)$. Thus, one and the same function $\Phi(e)$ serves as the indefinite integral for both $f(M)$ and $\varphi(M)$. This means that $f(M)$ and $\varphi(M)$ are equivalent. This proves the following theorem.

THEOREM 3. *The indefinite integral of a given summable function has the given function for its symmetric derivative almost everywhere.*

We agree to say that the point M_0 is a *Lebesgue point* of the summable function $f(M)$ if

$$\lim_{h\to 0}\frac{1}{h^2}\int\limits_{Q(M_0,h)}|f(M)-f(M_0)|\,dw=0.$$

THEOREM 4. *Almost all points of the rectangle R are Lebesgue points of any function $f(M)$ which is defined and summable on R.*

The proof of this theorem is entirely analogous to that of the one-dimensional case. Namely, we apply the preceding theorem to the function $|f(M)-r|$, where r is an arbitrary rational number. For almost all points $M_0\in R$, we have

$$\lim_{h\to 0}\frac{1}{h^2}\int\limits_{Q(M_0,h)}|f(M)-r|\,dw=|f(M_0)-r|.$$

Let $A(r)$ denote the set (of measure zero) of those points at which this relation is not satisfied and let

$$A=\sum A(r)+R(|f|=+\infty),$$

where the summation extends over all rational numbers. It is clear that $mA=0$.

We shall show that every point M_0 not in A will be a Lebesgue point of the function $f(M)$, which will prove the theorem.

Having taken $\varepsilon>0$, we choose an r such that

$$|f(M_0)-r|<\frac{\varepsilon}{3}. \tag{1}$$

Then

$$\Big|\,|f(M)-r|-|f(M)-f(M_0)|\,\Big|<\frac{\varepsilon}{3}$$

almost everywhere[8] and therefore

$$\left|\frac{1}{h^2}\int\limits_{Q(M_0,h)}|f(M)-r|\,dw-\frac{1}{h^2}\int\limits_{Q(M_0,h)}|f(M)-f(M_0)|\,dw\right|<\frac{\varepsilon}{3} \tag{2}$$

for every square $Q(M_0,h)$ contained in R. Since $M_0\overline{\in}A$, for every $\varepsilon>0$ selected, we can find a $\delta>0$ such that $0<h<\delta$ implies that

$$\left|\frac{1}{h^2}\int\limits_{Q(M_0,h)}|f(M)-r|\,dw-|f(M_0)-r|\right|<\frac{\varepsilon}{3}. \tag{3}$$

It follows from (1), (2), (3) that, for $0<h<\delta$,

$$\frac{1}{h^2}\int\limits_{Q(M_0,h)}|f(M)-f(M_0)|\,dw<\varepsilon,$$

which proves that M_0 is a Lebesgue point of the function $f(M)$.

---

[8] More precisely, this relation is true at all points M, where $f(M)\neq\infty$.

§3. GENERALIZATION OF THE PRECEDING RESULTS

We shall show in this section that, in the theorems introduced above, we could have discussed derivatives of a form more general than the symmetric derivative.

Let M be any point in the plane and let $\mathfrak{M}$ be a family of measurable sets, containing the point M. We agree to say that the family $\mathfrak{M}$ is *contractible*[9] to the point M if the family $\mathfrak{M}$ contains sets of arbitrarily small diameter. In this connection, the diameter dE of the set E is the least upper bound of the distances between all the points of E:

$$dE = \sup\{\rho(M, N)\}, \qquad (M \in E, N \in E).$$

If the family $\mathfrak{M}$ is contractible to the point M and if for every set e in $\mathfrak{M}$ a square $Q(M, h)$ containing it can be found such that

$$h^2 \leqslant \alpha \cdot me,$$

where α is a number which does not depend on the choice of the set e, then we say that the family $\mathfrak{M}$ contracts to the point M *regularly*. Roughly speaking, the regular contractibility of a family signifies that there are no sets in it which are too "flattened out."

Let $\mathfrak{M}$ be a family of measurable sets of positive measure, which is contractible to some point M, and let $\Phi(e)$ be an arbitrary set function. If the (finite or infinite) limit

$$\mathfrak{D}_M \Phi(e) = \lim_{de \to 0} \frac{\Phi(e)}{me}, \qquad (e \in \mathfrak{M})$$

exists, then this limit is called the *derivative* of the function $\Phi(e)$ at the point M, *with respect to the family* $\mathfrak{M}$.

THEOREM 1. *Let $\Phi(e)$ be an additive, absolutely continuous set function. At almost every point M of the fundamental rectangle R, the derivative $\mathfrak{D}_M \Phi(e)$ of this function exists with respect to an arbitrary family of sets, which is regularly contractible to this point, where the value of the indicated derivative does not depend on the choice of the family. For an arbitrary measurable set E contained in R, we have*[10]

$$\int_E \mathfrak{D}_M \Phi(e)\, dw = \Phi(E). \qquad (*)$$

Proof. The function $\Phi(e)$, being additive and absolutely continuous, is the indefinite integral of some summable function $f(M)$. Almost all points in R are Lebesgue points of this function. Let M_0 be a Lebesgue point of $f(M)$. We shall show that at this point the derivative $\mathfrak{D}_{M_0}\Phi(e)$, with respect to an arbitrary family, which is regularly contractible to M_0, equals $f(M_0)$:

$$\mathfrak{D}_{M_0} \Phi(e) = f(M_0).$$

This will then prove the theorem.

---

[9] This expression is used here for the first time.

[10] The derivative $\mathfrak{D}_M\Phi(e)$ may be defined not at all, but rather only at almost all, points of E. Therefore, we must understand the integral written in the last formula to be

$$\int_{E_0} \mathfrak{D}_M \Phi(e)\, dw,$$

where E_0 is that subset of E on which $\mathfrak{D}_M\Phi(e)$ is defined.

Let $\mathfrak{M}=\{e\}$ be any family of sets which is regularly contractible to the point M_0. Then for every set e of the family, we can find a square $Q(M_0, h)$ containing it, for which

$$h^2 \leqslant \alpha me.$$

Since

$$\frac{\Phi(e)}{me}-f(M_0)=\frac{1}{me}\int_e f(M)\,dw-f(M_0),$$

we have that

$$\left|\frac{\Phi(e)}{me}-f(M_0)\right| \leqslant \frac{1}{me}\int_e |f(M)-f(M_0)|\,dw.$$

But $Q(M_0, h)\supset e$. This means that

$$\int_e |f(M)-f(M_0)|\,dw \leqslant \int_{Q(M_0, h)} |f(M)-f(M_0)|\,dw.$$

If follows that

$$\left|\frac{\Phi(e)}{me}-f(M_0)\right| \leqslant \frac{\alpha}{h^2}\int_{Q(M_0, h)} |f(M)-f(M_0)|\,dw.$$

When $de\to 0$, then $h\to 0$ also; since M_0 is a Lebesgue point of the function $f(M)$, the last integral tends to zero as $h\to 0$. Thus,

$$\lim_{de\to 0}\frac{\Phi(e)}{me}=f(M_0).$$

This proves the theorem. It is useful to point out that it ceases to be valid if we omit the condition that the family of sets considered is regularly contractible.

To conclude this section, we stop to consider the question of how to obtain from the results of this chapter the results, already known, of Chapter IX, on functions of one variable.

Let $f(t)$ be a summable function on some interval (a, b). We obtain the indefinite integral of $f(x)$ by setting

$$\Phi(e)=\int_e f(t)\,dt, \qquad [e\subset(a, b)].$$

Its derivative with respect to an arbitrary family, which is regularly contractible to the point x, equals $f(x)$ at almost every point x. In particular, this is so if we take the family of segments $[x, x+\Delta x]$, having the point x as one of its endpoints,[11] as the family of sets considered. The derivative $\mathfrak{D}_x\Phi(e)$ with respect to this family coincides with the usual derivative, at the point x, of the function

$$\varphi(x)=\int_a^x f(t)\,dt,$$

because

$$\frac{\Phi([x, x+\Delta x])}{m[x, x+\Delta x]}=\frac{\varphi(x+\Delta x)-\varphi(x)}{\Delta x}.$$

[11]For this family, $\alpha = 2$.

Theorem 2 of §4, Chapt. IX is obtained again from this. In order to prove Theorem 3 of the same section we need the following theorem.

THEOREM 2. *If $\varphi(x)$ is an absolutely continuous function defined on the segment $[a,b]$, then there exists an additive, absolutely continuous function $\Phi(e)$, defined on the measurable sets of the segment $[a,b]$, such that*

$$\Phi([a,x]) = \varphi(x) - \varphi(a).$$

We shall give only a sketch of the proof of this theorem, leaving the details to the reader.

Since every absolutely continuous point function is the difference of two absolutely continuous increasing functions, we can assume without loss of generality that the function $\varphi(x)$ increases.

Let $a \leqslant \alpha \leqslant \beta \leqslant b$. Let $\mathscr{I}$ denote any one of the four intervals $[\alpha,\beta]$, (α,β), $[\alpha,\beta)$ and $(\alpha,\beta]$, and set

$$\Phi(\mathscr{I}) = \varphi(\beta) - \varphi(\alpha).$$

This defines the function $\Phi(e)$ *on all intervals lying in* $[a,b]$; we wish to define it *for every measurable set* $e \subset [a,b]$. The analogy between this problem and the problem of introducing the *measure* of a measurable set with which we were concerned in Chapter III is evident. This analogy also indicates a way to solve this problem. To wit: if G is an open set contained in $[a,b]$, then we set

$$\Phi(G) = \sum_{k=1}^{\infty} \Phi(\mathscr{I}_k),$$

where the $\mathscr{I}_k$ are the component intervals of G. It is not difficult to verify that $G_1 \subset G_2$ implies that $\Phi(G_1) \leqslant \Phi(G_2)$, that $G_1 G_2 = 0$ implies that $\Phi(G_1 + G_2) = \Phi(G_1) + \Phi(G_2)$, and that, in general, $\Phi(G_1 + G_2) \leqslant \Phi(G_1) + \Phi(G_2)$. All of this is established as in §1, Chapt. III.

Since the function $\varphi(x)$ is absolutely continuous, it is easily established that

$$\lim_{mG \to 0} \Phi(G) = 0. \qquad (*)$$

If E is an arbitrary measurable set contained in $[a,b]$, we remove the points a and b from it and consider all possible open sets G contained in (a,b) and containing E. We obtain the required function by setting

$$\Phi(E) = \inf\{\Phi(G)\}.$$

In fact, its absolute continuity is entirely obvious in virtue of relation (*). This means that the problem reduces to verifying the additivity of this function. But since the function $\Phi(e)$ has properties similar to those of the outer measure of the set e, then, as in Theorem 5, §3, Chapt. III, we obtain the inequality

$$\Phi(e_1 + e_2) \leqslant \Phi(e_1) + \Phi(e_2).$$

On the other hand, if F_1 and F_2 are two disjoint *closed* sets, we shall show, taking the separation theorem into consideration, that

$$\Phi(F_1 + F_2) = \Phi(F_1) + \Phi(F_2).$$

This is done exactly as in Theorem 6, §2, Chapt. III. Now we must establish the analogy between the function $\Phi(E)$ and the inner measure of the set E.

If E is an arbitrary measurable set and F is a closed subset of E, then $\Phi(F) \leqslant \Phi(E)$.

On the other hand, to every $\varepsilon > 0$ there corresponds a $\delta > 0$ such that the inequality $me < \delta$ implies the inequality $\Phi(e) < \varepsilon$. Having noted this, we take any measurable set E and $\varepsilon > 0$. We then find a δ corresponding to this ε (in the sense just indicated) and construct a closed set $F \subset E$ for which $mF > mE - \delta$.

Since $E = F + (E - F)$, we have

$$\Phi(E) \leqslant \Phi(F) + \Phi(E - F).$$

But $m(E - F) < \delta$ and therefore $\Phi(E - F) < \varepsilon$. Hence:

$$\Phi(E) < \Phi(F) + \varepsilon.$$

This shows that

$$\Phi(E) = \sup\{\Phi(F)\},$$

where the F are all possible closed sets contained in E. This proves that the function $\Phi(E)$ does in fact have properties similar to those of the inner measure. Taking this result into consideration and repeating the line of reasoning which was followed in detail in the proof of Theorem 6, §3, Chapt. III, we obtain that

$$\Phi(e_1 + e_2) \geqslant \Phi(e_1) + \Phi(e_2)$$

provided $e_1 e_2 = 0$.

Combined with the reverse inequality proved earlier, this relation establishes the additivity of the function $\Phi(e)$. This completes the proof of the theorem.

Now it is quite easy to prove Theorem 3, §4, Chapt. IX. To do this one must simply construct the function $\Phi(e)$, as was just pointed out. This function will be the indefinite integral of its derivative, i.e. $\Phi(e) = \int_e f(t)\,dt$.

In particular, for $e = [a, x]$, we obtain

$$\varphi(x) = \varphi(a) + \int_a^x f(t)\,dt,$$

which is equivalent to the theorem mentioned.

CHAPTER XIV

TRANSFINITE NUMBERS

§1. ORDERED SETS. ORDER TYPES

Beginning with Chapter III, we considered almost exclusively the metric theory of functions in which the basic concept is the *measure* of a point set. For this theory, the modest information from the theory of sets which was discussed in Chapters I and II was sufficient. Since we wish to consider certain questions in the descriptive theory of functions, we must now extend our information concerning set theory; the present chapter is devoted to this task.

In Chapter I, when we spoke of a set we abstracted from the *order* in which its elements were set down. Here, on the contrary, the question of the order of the elements of a given set will be fundamental.

DEFINITION 1. The set A is said to be *ordered* if a rule φ is given according to which of every two distinct elements a and b in the set A one precedes the other. In this connection, two requirements must be observed:

1) *If a precedes b, then b does not precede a;*

2) *If a precedes b and b precedes c, then a precedes c.*

If a precedes b, we say that b is a *successor* of a. Symbolically, we write this as

$$a \prec b, \quad b \succ a.$$

The rule φ itself, referred to in the definition, is called a *method of ordering*. When we say that an ordered set A is given, we shall always assume that it is being considered together with its method of ordering. If the sets

$$A = \{a, b, c\}, \quad B = \{b, c, a\}$$

are ordered in the lexicographical order, then A and B are *distinct* ordered sets.

In this connection, one finds the somewhat distinctive usage of the phrase "subset of an ordered set." Namely, if A is an ordered set and B is a subset of A, then any two elements of B are in A and have a definite order of succession there (in A). Hence, we agree to assume once for all that these elements have the same order of succession in B as in A. It is easy to see that this agreement is a method of ordering for the set B.

Thus, a *subset* of an ordered set A is an *ordered* set B, all elements of which are in A and whose mutual order is the same in B as in A.

For instance, if

$$A = \{a, b, c\}, \quad B = \{a, b\}, \quad C = \{b, a\},$$

then B is a subset of A, but C is not.

We now introduce some examples of ordered sets.

1. First of all, every set A of real numbers can be ordered by assuming that of two numbers in A, that one precedes which is the *smaller*. We shall call this the *natural* order.

2. We could also order any set of real numbers in the reverse order, assuming the *larger* number to be the predecessor. For example, the set of all natural numbers, ordered by this method, is written down as

$$\{\ldots, 5, 4, 3, 2, 1\}.$$

3. A finite set of n elements can be ordered by $n!$ distinct methods.

4. Every natural number n can be uniquely represented in the form

$$n = 2^k(2m+1) \qquad (k = 0, 1, 2, \ldots;\ m = 0, 1, 2, \ldots).$$

We agree to consider the number $n = 2^k(2m+1)$ the predecessor of the number $n' = 2^{k'}(2m'+1)$ if $k < k'$ or if $k = k'$, but $m < m'$.

With this method of ordering, the natural numbers are written down as

$$\begin{array}{cccccc} 1, & 3, & 5, & 7, & 9, & \ldots \\ 2, & 6, & 10, & 14, & 18, & \ldots \\ 4, & 12, & 20, & 28, & 36, & \ldots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{array}$$

where of two numbers in different rows that one precedes which is written higher and in every row the numbers are written down in their natural order. For instance,

$$7 \prec 12, \qquad 18 \prec 12, \qquad 28 \prec 36.$$

5. Every complex number can be written uniquely in the form

$$z = r(\cos\varphi + i\sin\varphi) \qquad (0 \leqslant \varphi < 2\pi)$$

where r is the modulus and φ is the argument of the number.[1] We order all complex numbers by agreeing that of two complex numbers that one precedes which has the smaller modulus, and in the case the moduli are equal, the one whose argument is smaller.

Suppose A is an ordered set and let $a \in A$. If there are no elements in the set A which precede a, then a is called the *first* element of A. The concept of *last* element is introduced analogously. Further, if a, b, c are elements of A and $a \prec b \prec c$, then we say that b lies *between* a and c.

DEFINITION 2. Suppose A and B are two ordered sets and let ψ be a 1-1 correspondence between them. If every relation

$$a \prec a',$$

which holds between elements of A, remains valid after replacing these elements by the elements of B corresponding to them, then the correspondence ψ is called a *similarity mapping of A onto B.*

In other words, a similarity mapping of one ordered set onto another ordered set is a one-to-one correspondence which leaves the order of succession of the elements invariant.

DEFINITION 3. If there exists a similarity mapping of an ordered set A onto the ordered set B, then we say the sets A and B are *similar* and we write this symbolically as

$$A \simeq B.$$

THEOREM 1. *Similar ordered sets are equivalent.*

In fact, a similarity mapping is a form of 1-1 correspondence.

---

[1] We agree to assume that $\varphi = 0$ for $z = 0$.

It is easy to see that for finite sets, this theorem is reversible, i.e. the following theorem is valid.

THEOREM 2. *If A and B are finite ordered sets consisting of the same number of elements, then these sets are similar.*

To prove this fact, we shall need the following lemma.

LEMMA. *If A is a finite ordered set, then it has a first element.*

In fact, take any element in A. If it is the first one, the lemma is proved. If not, there are elements in A preceding the one selected. Take any one of them. If it is the first element in A, the lemma is proved; otherwise, the possibility of further extraction of new elements from A presents itself. Continuing this process, we necessarily encounter the first element of A because otherwise we could extract an infinite sequence of distinct elements from a finite set, which is impossible.

COROLLARY. *If A is a finite ordered set, consisting of n elements, then it can be enumerated so that*

$$a_1 \prec a_2 \prec a_3 \prec \ldots \prec a_n.$$

In fact, to do this, one must simply denote the first element of A by a_1, the first element of the (finite) set $A - \{a_1\}$ by a_2, and so on.

Now Theorem 2 is clear. In fact, if A and B are ordered sets, each of which has n elements, then we enumerate them by the method indicated above and associate elements with the same indices.

Theorem 1 is not reversible for infinite sets. For example, if

$$A = \{1, 2, 3, \ldots\}, \qquad B = \{\ldots, 3, 2, 1\},$$

the sets A and B are equivalent (they consist of exactly the same elements). However, they are not similar, as is clear, say, from the fact that A has a first element but B does not, whereas under a similarity mapping of one ordered set onto another, the first element of one set must necessarily correspond to the first element of the other set.

THEOREM 3. *Let A, B, C be ordered sets. Then*

1) $A \simeq A$; 2) *If* $A \simeq B$, *then* $B \simeq A$; 3) *If* $A \simeq B$ *and* $B \simeq C$, *then* $A \simeq C$.

The proof is left to the reader.

Exactly as the concept of equivalence of two sets led us to the general definition of power, the concept of similarity leads to the definition of *order type*.

DEFINITION 4. Suppose the collection of all ordered sets is decomposed into classes, so that two sets are in the same class if, and only if, they are similar. We associate some symbol with every such class of ordered sets and we call it the *order type* of an arbitrary, representative, set of the given class.

The order type of an ordered set A is denoted by $\bar{A}$. If A and B are similar sets, we say that they have the same order type and write this as $\bar{A} = \bar{B}$.

All sets, having a given order type α, have the same power. It is called the power of the type α and is denoted by $\bar{\alpha}$. If $\bar{A} = \alpha$, then $\overline{\bar{A}} = \bar{\alpha}$.

If $\alpha = \beta$, then $\alpha = \bar{\beta}$, but as we saw above, the converse, generally speaking, is not true.

Some order types which are frequently encountered have generally accepted designations.

For example, the order type of the set

$$A = \{1, 2, \ldots, n\}$$

(and hence of every ordered set consisting of n elements) is denoted by the letter n. Thus, the symbol n serves to denote both the power and the order type of the set A. This imperfection in the notation does not lead to any noticeable inconvenience.

In the sequel, we shall agree to assume that the void set and sets consisting of one element are ordered and to denote their types by 0 and 1, respectively.

The order type of the set of all natural numbers, written down in the natural order

$$N = \{1, 2, 3, 4, \ldots\},$$

is denoted by the symbol ω, $\bar{N} = \omega$.

The order type of the set of all natural numbers, written down in the order inverse to the natural order,

$$N^* = \{\ldots, 4, 3, 2, 1\}$$

is denoted by the symbol ω^*.

Obviously $\omega^* = \bar{\omega}$, but $\omega^* \neq \omega$.

In general, if we have an ordered set A together with a method of ordering φ, we can obtain another ordered set A^* consisting of the same elements but with the "reverse" method of ordering, φ^*. Namely, if $a \in A$, $b \in A$ and $a \prec b$, according to φ, then according to φ^* we must have $a \succ b$. If α is the type of A, the type of A^* is denoted by α^*.

It is easily seen that $(\alpha^*)^* = \alpha$. For the order type n of a finite set, we have $n^* = n$.

The order type of the set of all integers in the natural order,

$$\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\},$$

is denoted by π. Obviously, $\pi^* = \pi$.

The order type of the set R of all rational numbers, written down in the natural order, is denoted by η. Clearly, $\eta^* = \eta$.

Finally, the order type of the set Z of all real numbers, in their natural order, is denoted by λ. It is clear that $\lambda^* = \lambda$. It is easy to see that the order type of an arbitrary open interval[2] (but not of a segment!) of the real line is λ. A similarity mapping of the interval $(a, b) = \{x\}$ onto the set $Z = \{y\}$ is defined, for example, by the formula

$$y = \tan \frac{(2x - a - b)\pi}{2(b - a)}.$$

THEOREM 4. *For any denumerable ordered set A, a subset[3] R_0 can be selected from the set R of all rational numbers, in their natural order, which is similar to the set A.*

Proof. Enumerate both the denumerable sets A and R:

$$A = \{a_1, a_2, a_3, \ldots\}, \qquad R = \{r_1, r_2 r_3, \ldots\}.$$

It is perfectly obvious that this enumeration is not connected in any way with the order of the elements in A and R (it is, of course, impossible to establish an enumeration of R for which $r_1 \prec r_2 \prec r_3 \prec \ldots$, because R does not contain a first element).

Set $n_1 = 1$. Having done this, denote by r_{n_2} that element in R which: 1) is written down with respect to r_{n_1}, exactly as the element a_2 is written down with respect to a_1 (i.e., if $a_2 \succ a_1$, then $r_{n_2} \succ r_{n_1}$, and if $a_2 \prec a_1$, then $r_{n_2} \prec r_{n_1}$); and 2) has the smallest index of all such elements in R. The existence of such an element r_{n_2} follows from the fact that R has neither a first nor a last element.

---

[2] We assume the numbers in the interval are arranged in the natural order.

[3] Here, the numbers in the set R_0 are arranged in R_0 in the same (natural) order as in R.

Further, making use of the fact that there are always intermediate elements between any two elements of R, and also that there is neither a first nor a last element in R, we can easily verify that there are elements in R written down with respect to r_{n_1} and r_{n_2} as a_3 is written down with respect to a_1 and a_2. We denote by r_{n_3} that one of them having the smallest index. Continuing this process indefinitely, we construct the sequence

$$r_{n_1}, r_{n_2}, r_{n_3}, \ldots$$

of elements in R, which is the required set R_0 (we note once more that the relative order of the numbers in $\{r_{n_k}\}$ is determined by their magnitude and not by the magnitude of their indices).

The theorem just proved can be formulated as follows: *a set of type η contains subsets of arbitrary type α and of power a.*

DEFINITION 5. Suppose A is an ordered set and let $a \in A$. The set of all elements in A, preceding the element a, is called the *segment of the set A determined by the element a* and is denoted by A_a.

We note that the element a itself is not in the segment A_a. If a is the first element in A, then A_a is the void set. A segment of an ordered set is a subset of the latter and, as such, is itself an ordered set.

Suppose $a \in A$ and let A_a be the segment of the set A determined by the element a. If a' is an element in A, preceding the element a, then the segments $A_{a'}$ and $(A_a)_{a'}$ of the sets A and A_a, respectively, determined by a', are identical:

$$(A_a)_{a'} = A_{a'},$$

so that of two segments of an ordered set one is a segment of the other. This makes it possible to order the set H consisting of all segments of an ordered set A. Namely, we agree to consider $A_{a'}$ to be the predecessor of two sets A_a and $A_{a'}$ if $A_{a'} \subset A_a$, or, equivalently, if $a' \prec a$. Obviously, the following theorem is valid.

THEOREM 5. *The set H of all segments of an ordered set A is similar to the set A.*

In fact, corresponding the element a with the segment A_a determined by it, we obtain a similarity mapping of the set A onto H.

For example, if $A = \{a, b, c\}$, then $H = \{0, \{a\}, \{a, b\}\}$. Both these sets have order type 3.

In conclusion, we stop to consider the operation of *adding order types.*

Assume $L = \{\lambda\}$ is an ordered set and let an ordered set A_λ be assigned to each $\lambda \in L$, where the sets A_λ are pairwise disjoint. In this case, it is easy to order the sum

$$S = \sum_{\lambda \in L} A_\lambda.$$

Namely, let a and a' be elements of S. Then

$$a \in A_\lambda, \quad a' \in A_{\lambda'}.$$

We agree to assume $a \prec a'$ (in the set S) if $\lambda \prec \lambda'$ (in the set L) or if $\lambda = \lambda'$, but $a \prec a'$ in the set A_λ. In the sequel, speaking of the sum of an ordered set of ordered sets, we shall always assume that this sum is ordered according to the method indicated above.

By what has been said above, the sets $A + B$ and $B + A$ are distinct ordered sets.

Now suppose $L = \{\lambda\}$ is an ordered set and let the order type α_λ be assigned to λ, for each λ. For each λ, we will take the set A_λ, ordered according to type α_λ and, assuming these sets are pairwise disjoint, we form the sum

$$S = \sum_{\lambda \in L} A_\lambda.$$

According to what was stated above, this sum is an ordered set. Its order type, by definition, is the sum $\bar{S}$ of the ordered set of order types $\{\alpha_\lambda\}$,

$$\bar{S} = \sum_{\lambda \in L} \alpha_\lambda.$$

It is evident that this definition does not depend on the choice of the sets A_λ but only on the types α_λ.

In simple cases, the order of the terms will be indicated by the very method of writing them down.

Examples.

1) $2 + 3 = 5$, because for $\bar{A} = 2$, $\bar{B} = 3$, we have $\bar{A} + \bar{B} = 5$. It is not difficult to verify that, in general, for a finite number of terms, which are types of finite sets, the new definition of sum coincides with what is usually understood by this term.

2) $1 + \omega = \omega$. In fact, $1 + \omega$ is the type of the set

$$\{a, b_1, b_2, b_3, \ldots\},$$

which has type ω.

3) In contrast, $\omega + 1 \neq \omega$, because $\omega + 1$ is the type of the set

$$\{b_1, b_2, b_2, \ldots; a\},$$

in which there is a last element. Thus, $1 + \omega \neq \omega + 1$ and therefore the addition of order types is not commutative.

4) $\omega^* + \omega = \pi$. Here, also $\omega + \omega^* \neq \omega^* + \omega$.

5) $1 + \lambda + 1$ is the order type of the segment $[a, b]$.

§2. WELL-ORDERED SETS

DEFINITION. If every non-void subset of an ordered set A has a first element, then A is said to be a *well-ordered set.*

Furthermore, we agree to consider the void set well-ordered.

THEOREM 1. *Every finite ordered set is well-ordered.*

In fact, every non-void subset of such a set is itself a finite ordered set and, by the lemma of §1, has a first element.

The following sets are other examples of well-ordered sets:

$$N = \{1, 2, 3, 4, \ldots\} \qquad (\bar{N} = \omega),$$
$$M = \{1, 3, 5, 7, \ldots; 2, 4, 6, 8, \ldots\} \qquad (\bar{M} = \omega + \omega).$$

As an example, we shall prove N is well-ordered. Let N^* be a non-void subset of N. Take an arbitrary element n in N^*. If n is the first element of N^*, then our proposition is proved. Otherwise, the set $N^* N_n$[4] is non-void and (together with N_n) it is a finite set; consequently, there is a first element n_0 in it which evidently is also the first element in N^*.

In contrast, the ordered set

$$L = \{\ldots, 4, 3, 2, 1\}$$

is not well-ordered.

The next theorem follows directly from the definition.

---

[4] N_n is the segment of N determined by the number n.

THEOREM 2. 1) *Every subset of a well-ordered set is well-ordered.*

2) *A non-void well-ordered set has a first element.*

3) *If of two similar ordered sets, one is well-ordered, then the other is also well-ordered.*

4) *Except for the last element, a well-ordered set has an immediate successor for every element.*

5) *It is impossible to select an infinite decreasing sequence, i.e. a sequence of the form*

$$a_1 \succ a_2 \succ a_3 \succ \dots, \tag{1}$$

from a well-ordered set.

We shall stop to prove 5) only. If the sequence (1) existed, then, being a (non-void) subset of a well-ordered set, it would contain a first element. But none of its elements a_n is the first because $a_{n+1} \prec a_n$.

The following theorem plays a fundamental role in the theory of transfinite numbers.

THEOREM 3. *Suppose A is a well-ordered set and let A^* be a subset of it (which may coincide with A). A similarity mapping[5] of the set A onto A^* in which the element a^* in A^* corresponds to the element $a \in A$ and precedes it (in the set A) cannot exist.*

Proof. We shall assume that, on the contrary, such similarity mappings of A onto A^* exist, and let φ be one of them. We denote by M the set of those elements in A to which there correspond, under the similarity mapping φ, elements in the set A^* preceding them (in A). By hypothesis, M is non-void and, hence, there is a first element a_0 in it. Let the element a_0^* in the set A^* correspond to the element a_0 under the similarity mapping φ. Obviously, in the set A, we have that

$$a_0^* \prec a_0.$$

But under the similarity mapping φ, to the element a_0^*, considered as an element of the set A, there also corresponds some element a_1 in the set A^*. Since φ is a similarity mapping, i.e. a correspondence leaving the order of succession of the elements in the set A^* invariant, and hence in the set A also, the element a_1 precedes the element a_0^*,

$$a_1 \prec a_0^*.$$

Thus, the element a_0^* must also be in M, which, obviously, is impossible because $a_0^* \prec a_0$ and a_0 is the first element in M. This completes the proof of the theorem.

COROLLARY 1. *A well-ordered set cannot be similar to any of its segments nor to a subset of any of its segments.*

In fact, if A is a well-ordered set and A_a is the segment determined by the element a, then for every similarity mapping of A onto A_a or onto a subset of A_a, to the element a there would have to correspond an element preceding it, which, as we saw, is impossible.

COROLLARY 2. *Two distinct segments of a well-ordered set cannot be similar.*

COROLLARY 3. *A well-ordered set cannot be similar to a segment of any of its subsets.*

THEOREM 4. *If A and B are two similar well-ordered sets, then there exists a unique similarity mapping of A onto B.*

Proof. Let φ and ψ be two distinct similarity mappings of the well-ordered set A onto the well-ordered set B. We denote by $\varphi(a)$ and $\psi(a)$, respectively, the images of the element $a \in A$ in the set B under these similarity mappings, and let a_0 be an element in A such that

---

[5]As usual, we assume that A^* is ordered in the same way as A, so that A^* is also well-ordered.

$$\varphi(a_0) \neq \psi(a_0)$$

(such an element exists inasmuch as φ and ψ are distinct similarity mappings). If $b = \varphi(a_0)$ and $b' = \psi(a_0)$, the segment A_{a_0} is similar to two distinct segments B_b and $B_{b'}$ of the set B. But then these segments must be similar, which, as we saw, is impossible.

The following theorem is fundamental in the theory of well-ordered sets.

THEOREM 5. *Of two well-ordered sets, one set is either similar to the other set or it is similar to a segment of the other set.*

Proof. Let A and B be two well-ordered sets. We agree to say that the element a of the set A is a "normal" element if the segment A_a determined by a is similar to some segment B_b of the set B. The first element of A can serve as an example of a normal element. It is evident that an element a', preceding a normal element a, is itself normal. In fact, if τ is a similarity mapping of the segment A_a onto the segment B_b similar to it, then obviously τ is a similarity mapping of the segment $A_{a'} = (A_a)_{a'}$ onto some segment of the segment B_b, i.e. onto a segment of the set B.

We denote the set of all normal elements of the set A by M. It turns out that either M coincides with the set A or is one of its segments. In fact, if $M \neq A$, then the set $(A - M)$ is non-void and there is a first element m in it. We shall show that

$$M = A_m. \tag{2}$$

If $a \in M$, then $a \neq m$. But it is also impossible to have $a \succ m$ because in this case the element m, which precedes the normal element a, would itself be normal and would lie in M, whereas $m \bar{\in} M$. Thus, it is necessary that $a \prec m$, and hence $a \in A_m$. This proves that $M \subset A_m$.

Conversely, if $a \in A_m$, then $a \prec m$ and, consequently, a cannot be in $A - M$; this means that $a \in M$. So $A_m \subset M$, and (2) is proved.

Now we shall call $b \in B$ a normal element if the segment B_b is similar to some segment of the set A; let N be the set of all normal elements in B. We show, in a manner analogous to that above, that $N = B$ or $N = B_n$.

We shall now show that the sets M and N are similar. Let $a \in M$. Then the segment A_a is similar to the segment B_b of the set B where, obviously, $b \in N$. We shall assume that the elements a and b correspond in a 1-1 manner. It is easy to see that to every element $a \in M$ there can correspond only one element b in the set N, and conversely (if two elements b and b' corresponded to the element a, then the segments B_b and $B_{b'}$, which are similar to the segment A_a, would be similar to one another, which is impossible).

We have thus established a 1-1 correspondence between the sets M and N. It remains to show that this correspondence leaves the order of the elements invariant, i.e. that it is a similarity mapping.

Let a and a' be two elements in M; let b and b' be the elements in N corresponding to them, and let $a \prec a'$. Suppose τ is a similarity mapping of the segment $A_{a'}$ onto the segment $B_{b'}$, which is similar to it; then τ is a similarity mapping of the segment $A_a = (A_{a'})_a$ onto some segment $(B_{b'})_{b_0}$ of the segment $B_{b'}$. But $(B_{b'})_{b_0} = B_{b_0}$ and, hence, the segment A_a is similar to the segment B_{b_0} of the segment B. On the other hand, there is only one segment in the set B which is similar to A_a, and this segment is B_b. Hence, $b = b_0$, and since $b_0 \in B_{b'}$, $b_0 \prec b'$; i.e. $b \prec b'$, which proves the similarity of the sets M and N.

Now it is easy to complete the proof. In fact, of the four logically possible cases

1) $M = A$, $N = B$, 3) $M = A$, $N = B_n$,
2) $M = A_m$, $N = B$, 4) $M = A_m$, $N = B_n$,

the fourth is impossible because it would mean that m is a normal element; but then $m \in M = A_m$, which contradicts the definition of a segment. So, the first three cases remain. In case 1), the sets A and B are similar; in cases 2) and 3) each of them is similar to a segment of the other. This completes the proof of the theorem.

If the well-ordered set A is similar to a segment of the well-ordered set B, we shall say the set A is *shorter* than the set B.

THEOREM 6. *A shortest well-ordered set exists in every set S of pairwise non-similar well-ordered sets.*

Proof. Let $A \in S$. If A is the shortest of the sets in S, the theorem is proved. Otherwise, S contains sets shorter than A and they are similar to some segments of A. Let $R = \{a\}$ be the set of elements a of the set A such that the segments A_a determined by them are similar to sets in S. If a^* is the first element in R and $A^* \in S$ is a set similar to the segment A_{a^*}, then A^* will be the shortest set in S. In fact, A^* is shorter than every set in B if A is shorter than B. If, however, B is shorter than A, then $B \simeq A_a$ for $a \in R$. But a^* is the first element in R so that $a^* \prec a$ and A_{a^*} is a segment of A_a, from which it is clear that A^* is shorter than B.

THEOREM 7. *The sum of a well-ordered collection of well-ordered sets is a well-ordered set.*

Proof. Let

$$S = \sum_{\lambda \in L} A_\lambda,$$

where the sets L and A_λ are well-ordered. The set S is ordered (see §1). Suppose S_0 is a non-void subset of S. We denote the set of those $\lambda \in L$ such that

$$A_\lambda S_0 \neq 0$$

by L_0 and let λ_0 be the first element in L_0. The set $A_{\lambda_0} S_0$ is a non-void subset of A_{λ_0}. If a_0 is the first element of this set, it is also the first element in the set S_0. This completes the proof of the theorem.

§3. ORDINAL NUMBERS

DEFINITION 1. The order type of a well-ordered set is called an *ordinal number*. If an ordinal number has power infinity, it is called a *transfinite* number.

The number 0 and all natural numbers are finite ordinal numbers. The numbers $\omega, \omega + 1, \omega + 2$ are transfinite numbers.

The order types $\omega^*, \pi, \eta, \lambda$ are not ordinal numbers because they are the types of ordered, but not well-ordered, sets.

DEFINITION 2. Let α and β be two ordinal numbers. We take two well-ordered sets A and B of type α and β, respectively. If A is shorter than B, we say α is *less than* β or that β is *greater than* α and write $\alpha < \beta$, $\beta > \alpha$.

This definition depends only on the numbers α and β and not on the sets A and B. In the application to finite numbers, the definition just stated is equivalent to the usual one so that

$$0 < 1 < 2 < 3 < \ldots,$$

and transfinite numbers are greater than all finite numbers.

An important characteristic of ordinal numbers is that *trichotomy* holds, i.e. the following theorem is valid.

THEOREM 1. *If α and β are ordinal numbers, each of the relations $\alpha = \beta$, $\alpha < \beta$, $\alpha > \beta$ excludes the others and one of them necessarily holds.*

In fact, let A and B be sets of type

$$\bar{A} = \alpha, \bar{B} = \beta.$$

If these sets are similar, $\alpha = \beta$. In addition, the relations $\alpha < \beta, \alpha > \beta$ are not possible because if $A \simeq B$, neither of these sets is shorter than the other. If the sets A and B are not similar, one (and only one) of them is necessarily shorter than the other.

Remark. If B is a subset of a well-ordered set A, then $\bar{B} \leqslant \bar{A}$.

In fact, by Corollary 3 to Theorem 3, §2, the relation $\bar{B} > \bar{A}$ is impossible.

THEOREM 2. *There is a smallest number in every set of pairwise unequal ordinal numbers.*

Proof. It is possible to assign a well-ordered set A of type $\bar{A} = \alpha$ to each $\alpha \in S$. If A_0 is the shortest of these sets (Theorem 6, §2), then $\alpha_0 = \bar{A}_0$ is the least of the numbers in the set S.

COROLLARY. *Every set of ordinal numbers, if ordered by the magnitude of its elements, is well-ordered.*

We agree to denote the set of all ordinal numbers less than the ordinal number α by W_α. According to the Corollary, W_α is a well-ordered set.

THEOREM 3. *The type of the set W_α is α,*

$$\bar{W}_\alpha = \alpha.$$

Proof. Let A be a set of type α. Denote the set of all segments of the set A by H. Then, by Theorem 5, §1, the type of H is α and it will suffice to show that H and W_α are similar.

Let A_a be an element of H. Its order type is a number less than α, i.e. an element in W_α. So, a definite element of W_α corresponds to each element of H. In this connection, distinct elements of W_α correspond to distinct elements of H because distinct elements of H are distinct segments of the set A, and these segments cannot be similar. Finally, every element of W_α is an ordinal number less than α; i.e. it is the type of some segment of the set A. Thus, we establish a 1-1 correspondence between H and W_α by assigning to elements of H their order types.

But this correspondence is a similarity mapping. In fact, suppose H is ordered so that of two segments of the set A which lie in H, we consider that one to be the first which is a segment of the other, i.e. whose type is less; this means that the relative order of two elements of H is the same as that of the numbers in W_α corresponding to them. This completes the proof of the theorem.

COROLLARY. *If A is a well-ordered set of type α, its elements can be enumerated using ordinal numbers less than α.*

In fact, the similarity mapping of the set A onto the set W_α assigns to each element of A an ordinal number less than α—i.e. its index. Then A can be written in the form of a sequence:

$$A = \{a_0, a_1, a_2, \ldots, a_\beta, \ldots\} \qquad (\beta < \alpha).$$

We note that the number 0 is in W_α, so that the first element in A has 0 for its index.

Finally, it will not be superfluous to mention that this similarity mapping of A onto W_α is possible in only one way.

The following, so-called Burali-Forti, antinomy is related to Theorem 3.

BURALI-FORTI ANTINOMY. Suppose W is the set of *all* ordinal numbers. By the corollary to Theorem 2, W is well-ordered. Let its order type be γ. Then W_γ has the same type. But W_γ is the segment of the set W determined by the element γ. Hence, the set W and its segment W_γ are similar, which contradicts Corollary 1 to Theorem 3, §2.

This antinomy shows that the very concept of the set of *all* ordinal numbers is intrinsically contradictory.

This sort of intrinsically contradictory concept occurs in set theory when the word "all" is used indiscriminately. We shall now attempt, for instance, to form the set S of *all* sets. Then S must contain itself as one of its elements, but this contradicts the agreement[6] that we always have $A \bar{\in} A$.

The reason why such contradictions appear lies in the fact that it is not in conformity with our rules to attempt to consider an object which occurs only in the process of its *formation* as something which is already *completed.*

Various mathematicians have suggested ways of constructing a theory of sets which enables one to avoid the occurrence of antinomies. With the "naive" treatment of the set concept which was adopted in the present book, it is impossible to go into detail in this connection. In the opinion of the author, one can characterize the aforementioned methods in general outline as various forms of the formalization of the idea, consisting in this, that the elements of any set, say, "precede" the concept of this set itself, and the formation of a set means the creation of a *new* mathematical object. It appears that antinomies ought not to appear upon the successive execution of this idea. For instance, if one carries out the aforementioned idea, then in forming the set S of *all* sets A, we can speak only about those sets A which are *distinct* from S, because until we introduced this S, it simply did not exist. But then the question of whether or not S contains itself as one of its elements, must be given a negative answer, and no violation of the rules of set theory occurs.[7]

Experience in set theory shows that in those cases where all the sets considered are subsets of some prescribed, completely determined, intrinsically non-contradictory[8] set, no absurd concepts occur. This observation also appears to be a confirmation of the point of view expounded above.

In conclusion, we note that the question of the admissibility of considering any set is connected with how correctly it is defined and not with whether it is finite or infinite: there exist examples of antinomies which appear in situations where no infinite sets

---

[6]Discarding this agreement is not the salvation of anything because *in this case an antinomy arises when we attempt to form the "set R of all sets which contain themselves." In fact, either of the assumptions* $R \in R$, $R \bar{\in} R$ *immediately leads to its contradiction.*

[7]We shall also assume that the agreement that we always have $A \bar{\in} A$ is a manifestation of the indicated idea.

[8]Unfortunately, there is no criterion enabling us to decide, with respect to every set, whether it is or is not a non-contradictory concept. The experience of the totality of our mathematical knowledge apparently furnishes the basis to assume that such "simple" sets as the set of natural numbers, the set of real numbers, and so on, represent sufficiently "trustworthy" creations.

occur. Thus, the antinomy problem is to a larger degree a problem of logic than of mathematics.

THEOREM 4. *The sum of a well-ordered set S of ordinal numbers is an ordinal number.*

This theorem follows from the definition of the sum of order types and Theorem 7, §2. It is self-evident that S is considered to be a "legitimate"[9] set; in particular

$$S \neq W.$$

THEOREM 5. *If $S=\{\alpha\}$ is a set of ordinal numbers, there exist ordinal numbers greater than all $\alpha \in S$.*

Proof. First of all we note that for every ordinal number α there exists a greater one, for example, $\alpha+1$. Therefore, the theorem is obvious for the case when S contains a largest number.

Let us now assume S has no largest number. Then the ordinal number

$$\sigma = \sum_{\alpha \in S} \alpha$$

is greater than all the numbers in S. In this connection, we assume that the set of elements in S is ordered according to their magnitude, so that S is well-ordered. To prove the inequality

$$\sigma > \alpha, \qquad (\alpha \in S)$$

we assign to each $\alpha \in S$ a well-ordered set A_α of type $\bar{A}_\alpha = \alpha$. Let

$$B = \sum_{\alpha \in S} A_\alpha \qquad (B = \sigma).$$

Then each set A_α is a subset of the segment B_b of the set B determined by the first element b of the set A_{α^*}, where $\alpha^* \in S$ and $\alpha^* > \alpha$. By Corollaries 1 and 3 to Theorem 3, §2, this means the set B is similar neither to A_α nor to the segment A_α, i.e. σ is not $= \alpha$ and σ is not $< \alpha$; it follows that $\sigma > \alpha$.

THEOREM 6. *The number $\alpha+1$ is the first number following α.*

Proof. Let β be the first number following α. Then

$$W_\beta = W_\alpha + \{\alpha\},$$

from which it follows, by the definition of the sum of order types, that

$$\beta = \overline{W}_\beta = \overline{W}_\alpha + \overline{\{\alpha\}} = \alpha + 1.$$

Whereas every number has a first successor, there exist numbers, for instance ω, which do not have a last predecessor. This motivates the following definition.

DEFINITION 3. An ordinal number is said to be a *number of the first* or *second type* depending on whether or not it has an immediately preceding number.

All finite numbers (except 0) are numbers of the first type; all numbers of the form $\alpha+1$, for example $\omega+1$, are also of this type. In contrast, ω is a number of the second type.

§4. TRANSFINITE INDUCTION

It is known what an important role the method of finite induction plays in mathematics. Let us recall it here.

THEOREM 1. *Let $T(n)$ be a proposition expressed in terms of the natural number n.*

[9] I. e. a set the consideration of which does not lead to any contradictions.

If

1) $T(n_0)$ *is true, and*

2) *the truth of* $T(n)$ *implies the truth of* $T(n+1)$,

then $T(n)$ *is true for all* $n \geqslant n_0$.

Proof. We assume there exist natural numbers $n \geqslant n_0$ such that $T(n)$ is false. Let n^* be the least such natural number. By 1), n^* is necessarily $> n_0$. Then $n^* - 1 \geqslant n_0$. It follows from the definition of n^* that $T(n^* - 1)$ is true; but then by 2) $T(n^*)$ is also true. This contradiction proves the theorem.

It is clear to the reader that the existence of the number n^* is guaranteed by the fact that the set of all natural numbers is well-ordered.[10] The same idea is used in the proof of the following *theorem on transfinite induction.*

THEOREM 2. *Let* $T(\alpha)$ *be a proposition expressed in terms of an ordinal number* α. *If*

1) $T(\alpha_0)$ *is true, and*

2) *the truth of* $T(\alpha)$ *for all* α, *such that* $\alpha_0 \leqslant \alpha < \beta$, *implies the truth of* $T(\beta)$,

then $T(\alpha)$ *is true for every ordinal number* $\alpha \geqslant \alpha_0$.

Proof. Assume $\alpha \geqslant \alpha_0$ exist for which $T(\alpha)$ is false. We denote the least such α by α^*. Then $\alpha^* > \alpha_0$ [because, by 1), $T(\alpha_0)$ is true] and for all α, where $\alpha_0 \leqslant \alpha < \alpha^*$, the proposition $T(\alpha)$ is true. It follows, by 2), that $T(\alpha^*)$ must be true. The contradiction thus obtained proves the theorem.

§5. THE SECOND NUMBER CLASS

DEFINITION. The set of all ordinal numbers which are types of denumerable well-ordered sets is called the *second number class* [11] K_0. [12]

It is clear that all numbers of the second class are transfinite.

THEOREM 1. *The number* ω *is the least number of the second class and, in general, it is the least transfinite number.*

Proof. By definition, ω is the type of the set

$$N = \{0, 1, 2, 3, 4, \ldots\}.$$

Every segment, N_n, of this set is a finite set. Hence an ordinal number less than ω is a finite number and, therefore, ω is the *smallest* transfinite number. Inasmuch as $\omega \in K_0$, the theorem is proved.

THEOREM 2. 1) *If* α *is a number of the second class, then* $\alpha + 1$ *is also a number of the second class.*

2) *If* S *is a denumerable set of numbers of the second class and* γ *is the smallest number greater than all the numbers of* S, *then* γ *is a number of the second class.*

Proof. The first part of the theorem is almost obvious. In fact,

---

[10] In a strictly formal discussion of the theory of natural numbers, Theorem 1 (or any proposition equivalent to it) is usually taken as an axiom. There is no concordance of opinion on the question to what extent this is necessary [see e.g. A. A. Markov, *Theory of Algorithms*, Mathematics Institute of the Academy of Sciences, USSR, 42, 1954, p. 18].

[11] The first number class is the set $N = \{0, 1, 2, 3, \ldots\}$.

[12] Here are some considerations whose purpose is to show that K_0 is a "legitimate" set. The set R of all rational numbers in their natural order evidently is a legitimate set. But then there is reason to believe that the set of all its subsets, which are ordered sets, is legitimate and, in particular, that the set of all well-ordered subsets of R, whose order types (Theorem 4, §1) comprise K_0, is legitimate.

$$\overline{\overline{\alpha+1}} = \overline{\overline{W_\alpha + \{\alpha\}}},$$

and the set $W_\alpha + \{\alpha\}$ is denumerable simultaneously with W_α.

Proceeding to the second part of the theorem, we may assume that S contains no greatest number, because otherwise we would be back at 1). But it is easy to show that under this assumption,

$$W_\gamma = \sum_{\alpha \in S} W_\alpha. \tag{1}$$

In fact, if any number occurs in the right member of (1), it obviously occurs in the left member. Conversely, if $\sigma \in W_\gamma$, then σ cannot be larger than all α in S and, consequently, there is a number α_0 in S such that $\alpha_0 \geqslant \sigma$. But α_0 is not the greatest number in S. Hence, a number $\alpha > \alpha_0$ can be found in S, and then $\sigma \in W_\alpha$.

The right member of (1) is a denumerable set; hence W_γ is denumerable and γ, being the order type of W_γ, occurs in K_0.

COROLLARY. *The class K_0 is non-denumerable.*

In fact, otherwise the first number following K_0 would occur in K_0.

The power of the set K_0 is denoted by the symbol $\aleph_1$ (read: aleph-one[13]), and the first number following K_0 is denoted by Ω.

THEOREM 3. *Powers intermediate to the power a of a denumerable set and the power $\aleph_1$ do not exist.*

Proof. The set W_Ω is the sum of a denumerable set $N = \{0, 1, 2, \ldots\}$ and K_0. This means that

$$\overline{\overline{W}}_\Omega = \aleph_1.$$

Noting this, we assume there exists a power m such that $a < m < \aleph_1$. Then a subset Q of power m can be extracted from W_Ω. The set Q is not similar[14] to W_Ω. Hence, Q is similar to a segment of W_Ω. But every segment of W_Ω is determined by a number from either N or K_0, i.e. it is a finite or a denumerable set. Therefore, Q is also finite or denumerable; this contradicts the assumption that $\overline{\overline{Q}} > a$.

THEOREM 4. *If $\alpha \in K_0$ is a number of the second type, then there exists an increasing sequence of ordinal numbers $\beta_1 < \beta_2 < \beta_3 < \ldots$, such that α is the smallest of the numbers greater than all numbers in this sequence.*

Proof. We enumerate the (obviously denumerable) set W_α of all numbers less than α: $\alpha_1, \alpha_2, \alpha_3, \ldots$.

There is no greatest number among the α_k (because otherwise α would be a number of the first type). Noting this, we set $n_1 = 1$ and denote the least natural number for which $\alpha_n > \alpha_{n_1}$ by n_2; then we denote the least natural number n for which $\alpha_n > \alpha_{n_2}$ by n_3, and so on. As a result, we obtain an increasing sequence $\alpha_{n_1} < \alpha_{n_2} < \alpha_{n_3} < \ldots$ with $n_1 < n_2 < n_3 < \ldots$.

We shall show this is the required sequence. It is clear that α is greater than all the numbers of the sequence and, consequently, it remains to show that no number less than α can be greater than all the numbers α_{n_k}.

Let $\gamma < \alpha$. Then $\gamma \in W_\alpha$ and, hence, $\gamma = \alpha_m$. If m coincides with one of the n_k, then γ is in $\{\alpha_{n_k}\}$, and therefore cannot be greater than all the numbers in $\{\alpha_{n_k}\}$. But if

[13]Aleph ($\aleph$) is the first letter of the Hebrew alphabet.

[14]Because Q is not even equivalent to W_Ω.

$n_k < m < n_{k+1}$, then, inasmuch as n_{k+1} is the smallest n for which $\alpha_n > \alpha_{n_k}$, it is clear that $\gamma < \alpha_{n_k}$. This completes the proof of the theorem.

In conclusion, we indicate the notation for certain numbers in K_0. We are already acquainted with the first of them — it is ω. The numbers following it, in order, are

$$\omega + 1, \omega + 2, \ldots, \omega + n, \ldots . \tag{2}$$

The first number following the numbers (2) will obviously be $\omega + \omega$; this number is usually denoted by $\omega \cdot 2$. The numbers following $\omega \cdot 2$ are

$$\omega \cdot 2 + 1, \ \omega \cdot 2 + 2, \ \ldots, \ \omega \cdot 2 + n, \ \ldots . \tag{3}$$

The first number following the numbers (3) is $\omega \cdot 3$.

Continuing this process, we define all numbers of the form $\omega \cdot n + m$.

There is a denumerable set of such numbers. The first number following them (and still belonging to K_0) is denoted by ω^2. The following numbers succeed ω^2:

$$\omega^2 + 1, \omega^2 + 2, \ldots, \omega^2 + n, \ldots;$$

the number $\omega^2 + \omega$ follows these numbers, and then the numbers

$$\omega^2 + \omega + 1, \omega^2 + \omega + 2, \ldots, \omega^2 + \omega + n, \ldots .$$

The number $\omega^2 + \omega \cdot 2$ follows these, and $\omega^2 + \omega \cdot 2$ is followed by $\omega^2 + \omega \cdot 2 + n$.

The first number following numbers of this form is the number $\omega^2 + \omega \cdot 3$. Numbers of the form $\omega^2 + \omega \cdot n + m$ are constructed in a similar manner.

The first number following them is $\omega^2 \cdot 2$, after which we introduce the sequence of numbers

$$\omega^2 \cdot 2 + \omega \cdot n + m.$$

After these numbers comes $\omega^2 \cdot 3$. This process leads us to the numbers

$$\omega^2 \cdot n + \omega \cdot m + l.$$

The first number following them is denoted by ω^3. After it come the numbers

$$\omega^3 + \omega^2 \cdot n + \omega \cdot m + l,$$

and then $\omega^3 \cdot 2$.

Continuing this process, we define the numbers ω^4, $\omega^5, \ldots$ and all "polynomials" of the form

$$\omega^k \cdot n + \omega^{k-1} \cdot n_1 + \ldots + \omega \cdot n_{k-1} + n_k. \tag{4}$$

The number ω^ω follows these numbers; this number is still in K_0 [because there is a denumerable set of numbers in (4)].

After ω^ω comes $\omega^\omega + 1$, and the entire process repeats itself from the beginning; this brings us to the number $\omega^\omega \cdot 2$. Then a new repetition of the process brings us to $\omega^\omega \cdot 3$. Thus, we construct all numbers of the form $\omega^\omega \cdot n$ The number $\omega^{\omega+1}$ follows them. Next we repeat the entire process from the beginning and we arrive at the number $\omega^{\omega+1} \cdot 2$. This introduces the numbers $\omega^{\omega+1} \cdot n$, and then $\omega^{\omega+2}$ follows them.

This method yields the numbers $\omega^{\omega+n}$ which are followed by the number $\omega^{\omega \cdot 2}$. The numbers $\omega^{\omega \cdot 2} + 1$, $\omega^{\omega \cdot 2} + 2, \ldots$ follow $\omega^{\omega \cdot 2}$, and everything is repeated from the beginning; this brings us to $\omega^{\omega \cdot 2} \cdot 2$.

Having defined $\omega^{\omega \cdot 2} \cdot n$, the first number following it is denoted by $\omega^{\omega \cdot 2 + 1}$. In a similar manner, greater and greater numbers are constructed,

$$\omega^{\omega \cdot 2 + n}, \ \omega^{\omega \cdot 3}, \ \omega^{\omega \cdot 3 + n}, \ \omega^{\omega \cdot 4}, \ \ldots, \ \omega^{\omega \cdot n}, \ldots$$

and after them, the number ω^{ω^2}.

The numbers ω^{ω^3}, ω^{ω^4}, ..., ω^{ω^ω}, ... are introduced similarly.

The first number following numbers of the form $\omega^{\omega^{\cdot^{\cdot^{\cdot^{\omega}}}}}$ is denoted by ε. The numbers following ε are $\varepsilon+1$, $\varepsilon+2$, and so on. Note, however, that we still do not have a notation for *all* numbers in K_0 because only a denumerable set of numbers are included in the above notation.

§6. ALEPHS

DEFINITION. The power of a well-ordered set is called an *aleph.*

All natural numbers (considered as a set) are alephs. The alephs of infinite sets are called *transfinite* alephs.

THEOREM 1. *Any two alephs are comparable.*

Proof. Let a and b be two alephs. We take the well-ordered sets A and B for which $\bar{\bar{A}}=a, \bar{\bar{B}}=b$.

If these sets are equivalent, then $a=b$. Otherwise, the sets are also not similar. But then one is shorter than the other. If A is shorter than B, then $a<b$.

The theorem is proved. It can also be formulated as follows: *if a and b are alephs, then one of the (pairwise incompatible) relations*

$$a=b, a<b, a>b$$

necessarily holds, i.e. trichotomy holds for the alephs.

THEOREM 2. *Let α and β be ordinal numbers and let $\bar{\alpha}$ and $\bar{\beta}$ be their alephs. Then*

1) $\bar{\alpha}<\bar{\beta}$ *implies* $\alpha<\beta$;

2) $\alpha<\beta$ *implies* $\bar{\alpha}\leqslant\bar{\beta}$.

Proof. The second part of the theorem is a consequence of the first. To prove the first part, we take the well-ordered sets A and B of type α and β, respectively. They are not similar because their alephs are not equal and they are also not equivalent. If B were shorter than A, this would mean, contrary to the hypothesis, that

$$\bar{\beta}=\bar{\bar{B}}\leqslant\bar{\bar{A}}=\bar{\alpha}.$$

Hence, A is shorter than B which in turn implies that $\alpha<\beta$.

It is easy to see that in the second part of the theorem, the equality sign must not be omitted, for example, $\omega<\omega+1$ but $\bar{\omega}=\overline{\omega+1}$.

THEOREM 3. *There is a least aleph in every set Q of pairwise unequal alephs.*

In fact, if we assign to each aleph in Q a well-ordered set which has this aleph for its power, we can guarantee the presence of a shortest set among these well-ordered sets. Its aleph will be the least in Q.

COROLLARY. *Every set of alephs, ordered according to their magnitude, is well-ordered.*

It is self-evident that we are dealing here with a "regular" set of alephs. It is impossible to consider the set of *all* alephs. This can be seen from the following theorem.

THEOREM 4. 1) *A greatest aleph does not exist.* 2) *For any set Q of alephs, there exist alephs greater than all alephs in Q.*

Proof. Let a be an aleph. This means there exists a well-ordered set A whose power is a. We consider, together with A, all well-ordered sets consisting of the same elements

but corresponding to other methods of ordering. The order types of these sets form some set T of ordinal numbers.

Let β be a number greater than all the numbers in T and let $b = \bar{\beta}$. Then b is an aleph and it can be shown that

$$b > a. \tag{1}$$

In fact, if $\alpha = \bar{A}$, then $\alpha \in T$ and $\alpha < \beta$. But then $a \leqslant b$ and it remains to disprove the possibility of the equality

$$b = a. \tag{2}$$

We assume this equality is true. Then it is possible to establish a one-to-one correspondence φ between A and W_β. Let A_0 be the set consisting of the same elements as A but ordered so that of any two of its elements, that one precedes which corresponds to the smaller number in the correspondence φ. The set A_0 is similar to W_β; this means its type is β and $\beta \in T$, which contradicts the definition of β. So, (2) is impossible and (1) is true.

Going over to the proof of the second part of the theorem, it is obvious we can assume there is no greatest aleph in Q and that the alephs in Q are pairwise unequal. We assign to each aleph in Q a well-ordered set A, having this aleph for its power, and we form the sum S of these sets, assuming the set of elements to be ordered in the same way as Q. But Q is well-ordered. Therefore, S, being the sum of a well-ordered set of well-ordered sets, is itself well-ordered and its power is aleph. It is clear that this aleph is greater than all alephs in Q. In fact, every aleph in Q is the power of a subset of S and hence it is not greater than $\bar{S}$; but if any of the alephs in Q coincided with $\bar{S}$, it would be the greatest in Q.

Remark. For every aleph, there is an aleph immediately following it. In fact, if a is an aleph, then there exists a greater aleph, b. If b follows immediately after a, our assertion is proved. Otherwise, we consider all[15] alephs m, for which $a < m < b$; there is a least one among them which is the one sought.

This remark enables us to establish a rational system of notation for all alephs. Namely, we denote by $\aleph_0$ the power of a denumerable set (it is clear this power is an aleph). The first succeeding aleph is denoted by[16] $\aleph_1$, the first following it by $\aleph_2$, and so on. The first aleph following all the alephs $\aleph_n$ is denoted by $\aleph_\omega$, and so on.

THEOREM 5. *Every aleph can be denoted by an $\aleph_\alpha$, where α is an ordinal number.*

Proof. Suppose H is an aleph and let T be the set consisting of all alephs less than H. The set T is well-ordered. Let its type be α. Let φ be a similarity mapping of T onto W_α; φ assigns to each aleph in T an order index less than α, after which it remains to denote H by $\aleph_\alpha$.

It is not difficult to verify that the system of notation is such that:

1) $\aleph_{\alpha+1}$ is the first aleph following $\aleph_\alpha$;

2) If the number β is the first following the numbers of the set $S = \{\alpha\}$, then $\aleph_\beta$ is the first aleph following all the alephs of the set $\{\aleph_\alpha\}$ $(\alpha \in S)$.

---

[15]The use here of the word "all" can be justified on the following basis: if b exists, then there exists a (legitimate) set B of power b. Its subsets form a legitimate set and, in particular, the set of those subsets of B, whose power m satisfies the inequality $a < m < b$, turns out to be legitimate.

[16]Above, we defined $\aleph_1$ as the power of the second number class K_0. The equivalence of the two definitions follows from Theorem 3, §5.

The set K_α of all ordinal numbers, having a given power $\aleph_\alpha$, is called a *number class*. The least of the numbers of the class K_α is usually denoted by Ω_α. In particular, $\Omega_0 = \omega$, $\Omega_1 = \Omega$. The class K_1 is the third number class.

§7. ZERMELO'S AXIOM AND THEOREM

The following proposition is used in many mathematical arguments.

ZERMELO'S AXIOM. *Let $S = \{M\}$ be a set of non-void and pairwise disjoint sets. Then there exists a set L possessing the following properties:*

1) $L \subset \sum_{M \in S} M$;

2) *the set L has one and only one element in common with each of the sets $M \in S$.*

One can say that L consists of "representatives" of all the sets $M \in S$.

The question whether this axiom is permissible gave rise to bitter controversies among mathematicians and there is no agreement on this point even at the present time. The author of this book takes the position of unqualified acknowledgment of Zermelo's axiom. In particular, we have already used it many times in the preceding chapters. Thus, for example, its use in Chapter III in the construction of an example of a non-measurable set is entirely obvious. But we also used Zermelo's axiom earlier. The existence of a denumerable subset in every infinite set, or the measurability of a (bounded) sum of a denumerable set of measurable sets are results which we established with the aid of Zermelo's axiom.[17]

For example, in proving Theorem 5, §3, Chapt. III, our line of reasoning was as follows: for every set E_k there is an entire family of open covering sets G_k. In order to form their sum $\sum G_k$, used in this theorem we, obviously, had to select from this family some one set G_k for each k, where this choice had to be carried out for all k at the same time. The possibility of such a choice could be based only on Zermelo's axiom. The matter is the same in the first of the cases mentioned.

Of course, Zermelo's axiom is not to be understood in the sense that we can *actually*[18] construct the set L. It is only a matter of the possibility of reasoning about this set without the risk of arriving at a contradiction.

The following theorem is closely related to Zermelo's axiom.

THEOREM 1 (GENERAL PRINCIPLE OF CHOICE). *Let $T = \{N\}$ be a collection of non-void sets. Then there exists a function $f(N)$, defined on T and which assigns to each set N in T a definite element of this set: $f(N) \in N$.*

Before proving this theorem, we note that in the case where the sets N are pairwise disjoint, the theorem is equivalent to Zermelo's axiom. In fact, applying Zermelo's axiom, we at once obtain the required function by setting $f(N) = NL$.

Conversely, applying Theorem 1, we obtain the set L by setting $L = \{f(N)\}$.

---

[17]It is recommended that the reader refer to the following articles concerning the Zermelo axiom: i) W. Sierpiński, *Zermelo's Axiom*, Mat. Sbornik, Vol. 31, 1921; ii) V. Molodshy, *Effectivism in Mathematics*, Sotsekgiz, 1938; iii) H. Lebesgue, *Integration and Finding Primitive Functions* (Supplement III, §2), GTTI, 1934; iv) H. Lebesgue, *Leçons sur l'intégration et la recherche des fonctions primitives*, Paris, 1928; v) W. Sierpiński, *Leçons sur les nombres transfinis*, Paris, 1928.

[18]The disagreement mentioned above is related to this: if the set L is not constructed, then it is impossible to decide for any object whether or not it occurs as an element in L. Therefore some scholars prefer not to consider L as a mathematical object at all.

We now proceed to the proof of the theorem. Denoting the elements of the sets N, which are in T, by n, we consider all pairs of the form

$$(n, N), \text{ where } n \in N \text{ and } N \in T.$$

In this connection, we shall consider the pairs (n', N') and (n'', N'') to be identical if and only if

$$n' = n'', \quad N' = N''.$$

We denote the set of all pairs (n, N_0) for which $n \in N_0$ by $M(N_0)$; such a set $M(N)$ can be constructed for every $N \in T$. Let

$$S = \{M(N)\} \quad (N \in T).$$

Distinct sets $M(N)$ are pairwise disjoint (the main idea of our line of reasoning lies in this fact). In fact, the situation that the sets $M(N')$ and $M(N'')$ are distinct implies that the sets N' and N'' are distinct. But then none of the pairs (n, N') coincides with any of the pairs (n, N'') and, hence

$$M(N') \cdot M(N'') = 0.$$

Applying Zermelo's axiom, we can ascertain the existence of a set L consisting of pairs (n, N), where $n \in N$, and having one element in common with each $M(N) \in S$. This set enables us to determine the required function. Namely, if $N_0 \in T$, the set $M(N_0)L$ consists of the single element (n_0, N_0), where $n_0 \in N_0$.

We satisfy the conditions of the theorem by setting $f(N_0) = n_0$.

COROLLARY. *Let $T = \{M'\}$ be the set of all non-void subsets of a given non-void set M. Then there exists a function $f(M')$ which assigns a definite element $m \in M'$ to each $M' \in T$.*

If we say that the element $m = f(M')$ is "marked" in M', then the corollary asserts that it is possible to mark one element[19] in each $M' \subset M$.

THEOREM 2 (E. ZERMELO). *Every set can be well-ordered.*

Proof. Suppose M is a non-void[20] set. We denote the set of all non-void subsets of M by $T = \{M'\}$ and "mark" one element in each M'.

Some subsets of M can be well-ordered. Such subsets are, for instance, all finite subsets of M. Let A be a non-void subset of M which is well-ordered by some method. If

(H) | *a is a marked element for every $a \in A$ in the subset $M - A_a$ of the set M (where A_a is the segment of A determined by the element a),*

then we shall say A is a *regularly ordered* subset of M.

Agreeing to denote the element marked in M' by $f(M')$, we can write condition (H) in the form:

$$f(M - A_a) = a \qquad (\text{for all } a \in A).$$

We first verify that regularly ordered subsets of M exist. In fact, if m_1 is a marked element in the set M itself and M_1 is the subset consisting of only that element, then M_1 is a regularly ordered subset because its only segment is the void set and

---

[19]Theorem 1 itself also implies that we can "mark" an element in each set $N \in T$ [$f(N)$ is such an element].

[20]The theorem is trivial for the void set.

$$f[M-(M_1)_{m_1}]=m_1.$$

Furthermore, if the marked element in the set $M-M_1$ is m_2, the ordered pair $\{m_1, m_2\}$ is a regularly ordered subset of M.

In every regularly ordered subset of A, the first element is necessarily the element m_1, which was marked in the set M itself. In fact, if a is the first element in A, then

$$a=f(M-A_a)=f(M)=m_1.$$

It is not difficult to show that a subset A of the set M can admit of only one method of regular ordering. In fact, let us assume that there exist two methods φ and ψ of ordering the subset A, as a result of each of which it becomes regularly ordered. Denote by P and Q those regularly ordered subsets into which A is transformed under the orderings φ and ψ.

One of these sets (say P) is either similar to the other or to a segment of it. Let τ be a similarity mapping of P onto Q (or onto a segment of it). In this connection, τ maps the element $m_1 \in P$ onto itself. If there were elements in P which were not mapped onto themselves, then one of them would be the first. Let such an element be p, which maps onto the element $q \neq p$. All elements preceding p map onto themselves. This means that $P_p = Q_q$.

However, this implies that

$$p=f(M-P_p)=f(M-Q_q)=q;$$

this contradicts the definition of p. Thus, all elements map onto themselves under the similarity mapping τ and this means that P and Q are identical.

Thus, if, in general, a non-void subset A of the set M admits of a regular ordering, this ordering is unique. We shall assume that every subset which can be regularly ordered has already been so ordered and we shall say it is *normal* under this hypothesis.

Of two distinct normal subsets, one is a segment of the other. In fact, if A and B are normal subsets, then one of them (say A) is either similar to the other or to a segment of it. If τ is a similarity mapping of A onto B (or onto a segment of B), then we can verify, as above, that τ maps every element of A onto itself. If A and B were similar, they would coincide. Hence, A is a segment of B.

It follows immediately from this assertion that if two elements of the set M occur in several normal subsets, their mutual order is the same in all of these subsets.

Having noted this, we denote the set of those elements of M which occur in at least one normal subset by L. The set L can be ordered. In fact, let a and b be any two elements in L. Then there exist normal subsets A and B such that $a \in A$ and $b \in B$. But one of these subsets certainly is contained in the other. Hence, both elements a and b occur in the same normal subset. In this subset, one of them precedes the other where, if $a \prec b$, they have this mutual order in every normal subset which contains both of them. We shall agree to assume that the mutual order of these elements in L is the same as in the normal subsets containing them.

Thus, the set L is ordered. But it is easy to verify that L is well-ordered. In fact, let L^* be a non-void subset of L. We take an arbitrary element a in L^*. If it is not the first element in L^*, then all the elements of L^* preceding a occur in every normal subset which contains a. Let A be such a subset. It is well-ordered and the intersection AL^* has a first element. This element is the first element in L^*.

We established that L is a well-ordered subset of M. We shall now show that it is regularly ordered. Let $a \in L$. Then we can find a normal subset A containing this element. It is entirely clear that $L_a = A_a$, from which it follows that $f(M - L_a) = f(M - A_a) = a$, and L is regularly ordered.

The proof is almost completed. Namely, it is not difficult to establish that $L = M$.

In fact, let us suppose this is not so, and let a be the marked element in the set $M - L$. We form the set $A = L + \{a\}$, assuming a follows all the elements in L. It is clear that this set is a normal subset of M. But then a must be in L, and at the same time we have $a \in M - L$.

This means that, in fact, $L = M$ and M is well-ordered. This completes the proof of the theorem.

COROLLARIES. 1. *Every power is an aleph.*

2. *Any two powers are comparable, i.e. trichotomy holds for powers.*

The first corollary implies that the power of the continuum c is an aleph,

$$c = \aleph_\alpha.$$

The *continuum* problem, as yet unsolved, consists in finding this α. The *continuum hypothesis*, which was mentioned in Chapter I, consists in this that $\alpha = 1$. Koenig,[21] a German mathematician, proved that $\alpha \neq \omega$. We note that the comparability of c and $\aleph_1$ and the inequality $c \geqslant \aleph_1$ follow from Theorem 4, §1 without Zermelo's theorem.[22]

---

[21]See I. I. Zhegalkin, *Transfinite Numbers*, Moscow, 1907, p. 337 or W. Sierpiński, *Leçons sur les nombres transfinis*, Paris, 1928.

[22]In fact, in the theorem referred to, we established that it is possible to extract $\aleph_1$ distinct ordered subsets from the set of all rational numbers R. But since, in general, there are c subsets, in toto, in R, we have that $\aleph_1 \leqslant c$.

CHAPTER XV

THE BAIRE CLASSIFICATION

§1. BAIRE CLASSES

We now consider the set of *all continuous* functions defined on a fixed segment $[a, b]$ and call this set the *null* class of functions. If the function $f(x)$, defined on the segment $[a, b]$, is not in the null class but is representable in the form

$$f(x) = \lim_{n\to\infty} f_n(x), \tag{1}$$

where all the functions $f_n(x)$ are continuous, then $f(x)$ is said to be a function of the *first* class.

In exactly the same way, a function $f(x)$ which is not in the null class nor in the first class, but is representable in the form (1), where all the functions $f_n(x)$ are of the first class, is said to be a function of the *second* class.

In general, the functions of class m are functions which are not in any of the preceding classes, but are representable as the limit of a sequence of functions of class $m-1$.

In this manner all classes, with finite indices, of functions are defined. We denote these classes by

$$H_0, H_1, \ldots, H_m, \ldots . \tag{2}$$

However, this classification can be continued still further. Namely, suppose $f(x)$ is not in any of the classes (2) and that it is representable in the form (1), where each of the functions $f_n(x)$ is in any of the classes H_{m_n}. Then we say this function is a function of class H_ω.

A function, which is not in any of the classes H_m nor in the class H_ω, but is the limit of a sequence of functions of class H_ω, is called a function of class $H_{\omega+1}$.

Let α be an ordinal number of the second number class. Assume we have already defined all classes H_β, where $\beta < \alpha$.

Then the class H_α is defined as the set of functions not in any of the classes $H_\beta\,(\beta < \alpha)$, but representable in the form (1), where every $f_n(x)$ belongs to some class $H_{\beta_n}\,(\beta_n < \alpha)$.

This classification is called the *Baire classification* and the functions of all classes $H_\alpha\,(\alpha < \Omega)$ are called *Baire functions.*

It is easy to see that the indices α of the classes H_α can only be numbers of the first or second number class and, in particular, it is impossible to define the class H_Ω by this method. In fact, no matter which denumerable sequence $\{f_n(x)\}$ of Baire functions we take, each of them falls into some class H_{α_n} for $\alpha_n < \Omega$:

$$f_n(x) \in H_{\alpha_n} \qquad (\alpha_n < \Omega).$$

But then there exists a number γ, which is greater than all the numbers α_n

($n = 1, 2, 3, \ldots$) and which, in addition, belongs to the second class. This means that if the sequence $\{f_n(x)\}$ converges to some function $f(x)$, then the latter function is in the class whose index is not greater than the number γ.

Further, it is not difficult to show that if α is a number of the first type, i.e. $\alpha = \beta + 1$, then every function $f(x) \in H_\alpha$ is representable in the form (1), where all the functions $f_n(x)$ are in the class H_β. In fact, if there were an infinite set of functions $f_n(x)$ of class H_γ for $\gamma < \beta$, then $f(x)$ itself would be in H_β or in a class with still smaller index. Therefore, all functions $f_n(x)$, beginning with some one of them, are in H_β.

But if $f(x) \in H_\alpha$, where α is a number of the second type, then $f(x)$ admits of a representation (1), in which

$$f_n(x) \in H_{\beta_n} \qquad (\beta_1 < \beta_2 < \beta_3 < \ldots < \alpha). \tag{3}$$

In fact, let $f_1(x) \in H_{\beta_1}$. Among the functions $f_2(x), f_3(x), \ldots$ one can necessarily be found which belongs to the class H_{β_2} with $\beta_2 > \beta_1$ (otherwise, $f(x) \in H_{\beta_1 + 1}$). Continuing this process, we select a subsequence from $\{f_n(x)\}$ which satisfies (3).

THEOREM 1. *All Baire functions are measurable.*[1]

This assertion follows from the fact that continuous functions are measurable, and taking the limit does not take us out of the class of measurable functions.

The converse statement is not true as is shown by the next theorem.

THEOREM 2. *The set of all Baire functions has power c.*

Proof. The power of the set of all continuous functions is c. This means there are c functions in the class H_0. We shall prove that

$$\bar{\bar{H}}_\alpha \leqslant c \tag{4}$$

for every α.

This is true for $\alpha = 0$. Assume (4) is proved for all $\alpha < \beta$, where β is some number of the first or second class; we shall show that (4) is also true for $\alpha = \beta$.

To this end, set

$$T_\beta = \sum_{\xi < \beta} H_\xi.$$

Each H_ξ has power not greater than c and there are a finite or a denumerable number of terms in the sum under consideration. This means that $T_\beta \leqslant c$. On the other hand, $H_0 \subset T_\beta$, from which it follows that $\bar{\bar{T}}_\beta \geqslant c$, and hence

$$\bar{\bar{T}}_\beta = c. \tag{5}$$

Having noted this, we now consider the set

$$M = \{(f_1(x), f_2(x), f_3(x), \ldots)\}$$

of all possible sequences of functions from the set T_β. The elements of the set M are determined by a denumerable set of parameters, each of which assumes c values, according to (5). Therefore the power of M equals c (cf. Chapt. I, §4, Theorem 7).

But it is possible to assign a definite element of M to every function $f(x)$ of H_β, namely, that one for which

$$\lim_{n \to \infty} f_n(x) = f(x).$$

[1]Moreover, it can be shown that a necessary and sufficient condition that $f(x)$, defined on $[a, b]$, be a Baire function is that it be (B)-measurable.

This means that H_β is equivalent to a subset of M, and this proves (4) for $\alpha = \beta$. By the theorem on transfinite induction, (4) is then proved for all $\alpha < \Omega$.

Now it is already easy to complete the proof. Namely, if T is the set of all Baire functions, then

$$T = \sum_{\alpha<\Omega} H_\alpha.$$

Each term has power not greater than c, and the set of terms has power $\aleph_1$, which is also not greater than c; it follows that $\bar{\bar{T}} \leqslant c$. But since $\bar{\bar{T}} \geqslant \bar{\bar{H}}_0 = c$, we have $\bar{\bar{T}} = c$, and the theorem is proved.

Let A and B be two real numbers, where $A < B$. We shall agree to denote the following function of x by the symbol $[x]_A^B$:

$$[x]_A^B = \begin{cases} B \text{ if } x > B, \\ x \text{ if } A \leqslant x \leqslant B, \\ A \text{ if } x < A. \end{cases}$$

LEMMA 1. *If*

$$\lim_{n\to\infty} x_n = l,$$

then

$$\lim_{n\to\infty} [x_n]_A^B = [l]_A^B.$$

Proof. First of all, let $l \neq A$ and $l \neq B$, for instance, $A < l < B$. Then for sufficiently large n we have $A < x_n < B$, so that

$$[x_n]_A^B = x_n \to l = [l]_A^B.$$

The cases $l < A$ and $l > B$ are investigated analogously.

Now we assume that $l = B$. Taking an arbitrary $\varepsilon > 0$, we find an N such that for $n > N$,

$$x_n > B - \varepsilon, x_n > A.$$

Let $n > N$. One of the following is true: either $x_n \leqslant B$ and then $[x_n]_A^B = x_n$, or $x_n > B$ and then $[x_n]_A^B = B$. In both cases we have

$$B - \varepsilon < [x_n]_A^B \leqslant B,$$

from which it follows that

$$\lim_{n\to\infty} [x_n]_A^B = B.$$

For $l = A$, the reasoning is analogous.

COROLLARIES. 1. *If $f(x)$ is a continuous function, then $[f(x)]_A^B$ is also a continuous function.*

2. *If $f(x)$ is a function of a class with index $\leqslant \alpha$, then $[f(x)]_A^B$ is also a function of a class with index $\leqslant \alpha$.*

3. *If $f(x) \in H_\alpha$ and $A \leqslant f(x) \leqslant B$, then $f(x)$ can be written as the limit of a sequence of functions $f_n(x)$, each of which is in a class H_{α_n} $(\alpha_n < \alpha)$ and satisfies the inequality $A \leqslant f_n(x) \leqslant B$.*

The first and third corollaries are obvious, and the second is proved by the method of transfinite induction.

LEMMA 2. *If*

$$\lim_{n\to\infty} x_n = l,$$

then also

$$\lim_{n\to\infty} [x_n]^n_{-n} = l.$$

The proof is left to the reader.

COROLLARY. *Every function $f(x)$ of the class H_α is representable in the form*

$$f(x) = \lim_{n\to\infty} f_n(x),$$

where the $f_n(x)$ are bounded functions of the classes H_β, for $\beta < \alpha$.

The functions can, a fortiori, be considered *finite*.

THEOREM 3. *The sum, difference, and product of two finite functions of classes $\leqslant \alpha$ is a function of class $\leqslant \alpha$. The same is also true for the quotient, provided the function in the denominator does not vanish.*

Proof. Let $f(x)$ and $g(x)$ be finite functions each of which is in some class of index $\leqslant \alpha$, and let $s(x) = f(x) + g(x)$.

We shall show that $s(x)$ is also a function of class $\leqslant \alpha$. If $\alpha = 0$, this assertion is trivial. We shall assume that it is proved for all $\alpha < \lambda$ and show it is also true for $\alpha = \lambda$.

To this end, we shall construct two sequences of finite functions $\{f_n(x)\}$ and $\{g_n(x)\}$ for which

$$\lim_{n\to\infty} f_n(x) = f(x), \quad \lim_{n\to\infty} g_n(x) = g(x)$$

and

$$f_n(x) \in H_{\beta_n}, \quad g_n(x) \in H_{\gamma_n} \qquad (\beta_n < \lambda, \ \gamma_n < \lambda).$$

We set

$$s_n(x) = f_n(x) + g_n(x).$$

Then, by hypothesis, $s_n(x) \in H_{\lambda_n}$, where λ_n is less than or equal to the larger of the numbers β_n and γ_n, and hence $\lambda_n < \lambda$. But $\lim s_n(x) = s(x)$, from which it is clear that $s(x)$ is in a class with index $\leqslant \lambda$.

For the difference and product, the reasoning is entirely analogous, and for the quotient, it is necessary to introduce the function

$$\frac{f_n(x)\, g_n(x)}{g_n^2(x) + \dfrac{1}{n}}.$$

LEMMA 3. *Let*

$$A_1 + A_2 + A_3 + \dots$$

be a given convergent numerical series with positive terms. If each of the functions in the sequence $\{f_k(x)\}$ is a function of class $\leqslant \alpha$ $(\alpha > 0)$ and satisfies the inequality

$$|f_k(x)| \leqslant A_k \qquad (k = 1, 2, 3, \dots),$$

then the sum of the series $\sum_{k=1}^{\infty} f_k(x)$ is a function of class $\leqslant \alpha$.

Proof. Each of the functions $f_k(x)$ can be written in the form

$$f_k(x) = \lim_{n\to\infty} \varphi_n^{(k)}(x),$$

where $\varphi_n^{(k)}(x)$ is in a class with index *less than* α, and

$$|\varphi_n^{(k)}(x)| \leqslant A_k \qquad (n = 1, 2, 3, \dots).$$

We set

$$\Phi_n(x) = \varphi_n^{(1)}(x) + \varphi_n^{(2)}(x) + \ldots + \varphi_n^{(n)}(x).$$

This function is in a class with index $< \alpha$, and to prove the lemma, it is sufficient to verify that

$$f(x) = \lim_{n \to \infty} \Phi_n(x),$$

where $f(x)$ denotes the sum of the series $\sum f_k(x)$.

For this purpose, we take $\varepsilon > 0$ and find an m such that

$$\sum_{k=m+1}^{\infty} A_k < \varepsilon.$$

Then

$$\left| \sum_{k=m+1}^{\infty} f_k(x) \right| < \varepsilon, \quad \left| \sum_{k=m+1}^{n} \varphi_n^{(k)}(x) \right| < \varepsilon \qquad (n > m),$$

and therefore

$$|f(x) - \Phi_n(x)| < \sum_{k=1}^{m} |f_k(x) - \varphi_n^{(k)}(x)| + 2\varepsilon.$$

But for every fixed x, the difference $f_k(x) - \varphi_n^{(k)}(x)$ tends to zero with increasing n. Therefore, for $n > N(x)$, we have

$$|f(x) - \Phi_n(x)| < 3\varepsilon,$$

which proves the lemma.

THEOREM 4. *The limit of a uniformly convergent sequence of functions of classes* $\leqslant \alpha$ *is a function of class* $\leqslant \alpha$.

Proof. For $\alpha = 0$, the theorem is trivial. We shall consider the case when $\alpha > 0$. Let

$$f(x) = \lim_{n \to \infty} f_n(x),$$

where the class of each of the functions $f_n(x)$ is not greater than α, and passage to the limit is uniform with respect to x.

We select a sequence of indices $n_1 < n_2 < n_3 < \ldots$ so that

$$|f_{n_k}(x) - f(x)| < \frac{1}{2^k} \qquad (a \leqslant x \leqslant b).$$

Then the terms of the series

$$[f_{n_2}(x) - f_{n_1}(x)] + [f_{n_3}(x) - f_{n_2}(x)] + [f_{n_4}(x) - f_{n_3}(x)] + \ldots \tag{6}$$

are not greater than the terms of the convergent series with positive terms

$$1 + \tfrac{1}{2} + \tfrac{1}{4} + \ldots,$$

and therefore, by the lemma, the sum of the series (6), which is evidently equal to

$$f(x) - f_{n_1}(x), \tag{7}$$

is a function of class not greater than α. The function $f(x)$ lies in a class of index $\leqslant \alpha$ together with the difference (7).

Generally speaking, the operation of taking the limit leads to functions of a higher class than the classes to which the functions of the sequence belong. We saw that the condition of uniform convergence was *sufficient* that there be no increase in the class.

B. M. Gagayev[2] found *necessary and sufficient conditions* which must be put upon the functions of the sequence in order that the class of the limit function not be greater than the class containing the functions of the sequence.

THEOREM 5. *Let $f(x)$ be a function of class not greater than β, and let $\varphi(t)$ be a function of class not greater than α, whose values lie in the segment $[a, b]$ on which $f(x)$ is defined. Then the composite function $f[\varphi(t)]$ is a function of class $\leqslant \alpha+\beta$.*

Proof. Let $\beta=0$, i.e. let $f(x)$ be continuous. We shall show that the function $f[\varphi(t)]$ will be of class $\leqslant\alpha$ if $\varphi(t)$ is of class $\leqslant\alpha$.

For $\alpha=0$, this assertion is trivial. If it is proved for all $\alpha<\gamma$ and $\varphi(t)\in H_\gamma$, then

$$\varphi(t)=\lim_{n\to\infty}\varphi_n(t),$$

where $\varphi_n(t)$ is in the class $H_{\gamma_n}(\gamma_n<\gamma)$ and it can be assumed that

$$a\leqslant\varphi_n(t)\leqslant b.$$

But then, by virtue of the continuity of $f(x)$,

$$f[\varphi(t)]=\lim_{n\to\infty}f[\varphi_n(t)]$$

and, inasmuch as the class of $f[\varphi_n(t)]$ is not greater than γ_n, the class of $f[\varphi(t)]$ is not greater than γ.

So, if $\varphi(t)$ is a function of class $\leqslant\alpha$, then for every function $f(x)$ of class β, the composite function $f[\varphi(t)]$ is of class $\leqslant\alpha+\beta$ with $\beta=0$.

Let us assume this assertion has already been proved for all $\beta<\gamma$ and let $f(x)\in H_\gamma$.

Then

$$f(x)=\lim_{n\to\infty}f_n(x),$$

where the class of $f_n(x)$ is H_{γ_n} for $\gamma_n<\gamma$. This means that

$$f[\varphi(t)]=\lim_{n\to\infty}f_n[\varphi(t)],$$

and the class of $f_n[\varphi(t)]$ is not greater than $\alpha+\gamma_n$, i.e. less than[3] $\alpha+\gamma$. This implies that the class of $f[\varphi(t)]$ is not greater than $\alpha+\gamma$. By the principle of transfinite induction, the theorem is proved.

In conclusion, we note that all definitions and theorems introduced above carry over word for word to the case of functions of several variables, defined on some parallelepiped.

For example, Theorem 5 takes the following form.

THEOREM 5*. *Let $f(x_1, x_2, \ldots, x_n)$ be a function of class $\leqslant\beta$, defined in the parallelepiped $a_k\leqslant x_k\leqslant b_k (k=1,2,\ldots,n)$, and let $\varphi_1(t_1,\ldots,t_m)$, $\varphi_2(t_1,\ldots,t_m)$, $\ldots$, $\varphi_n(t_1,\ldots,t_m)$ be a family of n functions of classes $H_{\alpha_1}, H_{\alpha_2},\ldots,H_{\alpha_n}$, respectively, such that $a_k\leqslant\varphi_k(t_1,\ldots,t_m)\leqslant b_k$. Then the composite function $f(\varphi_1,\varphi_2,\ldots,\varphi_n)$ is in a class of index $\leqslant\alpha+\beta$, where α is the greatest of the numbers $\alpha_1,\ldots,\alpha_n$.*

§2. NON-VACUOUSNESS OF THE BAIRE CLASSES

The natural question arises whether there exist functions of class H_α for every $\alpha<\Omega$. Following H. Lebesgue, we shall show that this is actually so. To do this, we

[2] *Sur les suites convergentes de fonctions mesurables B*, Fund. Math., Vol. 18, 1931, pp. 182-188.

[3] We use the evident fact that $\alpha+\sigma<\alpha+\gamma$ for $\sigma<\gamma$ (the order of the terms is essential here!).

need the concepts, important in themselves, of the limit superior and limit inferior of a sequence of numbers.

Let

$$x_1, x_2, x_3, \ldots \tag{1}$$

be a numerical sequence. We set[4]

$$\bar{x}_n = \sup\{x_n, x_{n+1}, x_{n+2}, \ldots\}.$$

It is easy to see that

$$\bar{x}_1 \geqslant \bar{x}_2 \geqslant \bar{x}_3 \geqslant \ldots,$$

and, therefore, there exists a definite (finite or infinite) limit

$$\lim_{n\to\infty} \bar{x}_n.$$

This limit is called the *limit superior* of the sequence (1) and is denoted by

$$\overline{\lim_{n\to\infty}}\, x_n.$$

Analogously, the limit

$$\lim_{n\to\infty} \underline{x}_n,$$

where $\underline{x}_n = \inf\{x_n, x_{n+1}, x_{n+2}, \ldots\}$, is called the *limit inferior* of the sequence (1) and is denoted by

$$\underline{\lim_{n\to\infty}}\, x_n.$$

It follows from the obvious inequality $\underline{x}_n \leqslant \bar{x}_n$ that

$$\underline{\lim}\, x_n \leqslant \overline{\lim}\, x_n.$$

THEOREM 1. *If* $b = \overline{\lim}\, x_n$, *then the sequence* $\{x_n\}$ *contains a subsequence which converges to* b.

Proof. If $b = -\infty$, the theorem is trivial, because, as follows from the inequality $x_n \leqslant \bar{x}_n$, in this case the sequence $\{x_n\}$ itself tends to $-\infty$. Further, let $-\infty < b < +\infty$. Then a k_0 can be found such that for all $k > k_0$, the numbers x_k will also be finite. Assign an m_k to each $k > k_0$ such that

$$\bar{x}_k - \frac{1}{k} < x_{m_k} \leqslant \bar{x}_k \qquad (m_k \geqslant k).$$

Obviously,

$$\lim_{k\to\infty} x_{m_k} = b.$$

Since we wish to obtain a sequence $\{x_{n_i}\}$ converging to b and such that $n_1 < n_2 < n_3 < \ldots$, we set $n_1 = m_1$ and then denote the smallest of the numbers m_k which are greater than n_1 by n_2; the smallest of the m_k greater than n_2 by n_3; and so on. Inasmuch as the sequence $\{x_{n_i}\}$ will be a subsequence of $\{x_{m_k}\}$, it also converges to b.

It remains to consider the case when $b = +\infty$. In this case, for all $k, \bar{x}_k = +\infty$, i.e. none of the sets $\{x_k, x_{k+1}, x_{k+2}, \ldots\}$ is bounded above. We set $n_1 = 1$ and let n_{i+1} be chosen so that the conditions

$$x_{n_{i+1}} > x_{n_i} + 1, \quad n_{i+1} > n_i$$

---

[4]It is possible that $\bar{x}_n = +\infty$, and also that $\bar{x}_n = -\infty$ (the last relation means that $x_n = x_{n+1} = \ldots = -\infty$, which we do not exclude).

are satisfied. This at once leads to the required subsequence. The theorem is proved.

If $b^* > b$, an $\bar{x}_{n_0}$ can be found such that $\bar{x}_{n_0} < b^*$. For $n \geqslant n_0$,

$$x_n \leqslant \bar{x}_{n_0} < b^*,$$

and the number b^* cannot be the limit of any subsequence in $\{x_n\}$. Thus, the limit superior of a numerical sequence can be defined as the greatest of the numbers which are limits of subsequences of the given sequence. An analogous remark can be made with respect to the limit inferior.

THEOREM 2. *If the sequence* (1) *has a (finite or infinite) limit l, then*

$$\underline{\lim}\, x_n = \overline{\lim}\, x_n = l.$$

Conversely, if the limit superior and limit inferior of the sequence (1) *are equal, their common value is the limit of the sequence.*

Proof. If sequence (1) has a limit l, then all its subsequences also converge to the same limit; this implies the first part of the theorem.

The second part is an immediate consequence of the obvious inequality

$$\underline{x}_n \leqslant x_n \leqslant \bar{x}_n.$$

We introduce the following notation. Let $a, b, \ldots, l$ be a finite system of numbers. We denote the largest of them by $\max\{a, b, \ldots, l\}$.

LEMMA 1. *If*

$$x_n \to a, y_n \to b, \ldots, z_n \to l,$$

then

$$\max\{x_n, y_n, \ldots, z_n\} \to \max\{a, b, \ldots, l\}.$$

We shall leave the proof to the reader.

COROLLARIES. 1. *If the functions $f_1(x), f_2(x), \ldots, f_n(x)$ are continuous, then the function*

$$\varphi(x) = \max\{f_1(x), f_2(x), \ldots, f_n(x)\} \tag{2}$$

is also continuous.

2. *If the functions $f_1(x), \ldots, f_n(x)$ are in classes with indices $\leqslant \alpha$, then function* (2) *is also of class $\leqslant \alpha$.*

The first corollary is evident; the second is proved by the method of transfinite induction.

LEMMA 2. *Let $\{x_n\}$ be a numerical sequence. If*

$$l = \sup\{x_n\}, \quad y_n = \max\{x_1, x_2, \ldots, x_n\},$$

then

$$l = \lim_{n \to \infty} y_n.$$

The proof is left to the reader.

THEOREM 3. *Let*

$$f(x) = \overline{\lim_{n \to \infty}}\, f_n(x).$$

If each $f_n(x)$ is a function of class $\leqslant \alpha$, then $f(x)$ is a function of class $\leqslant \alpha + 2$.

Proof. Let

$$\bar{f}_{n,p}(x) = \max\{f_n(x), f_{n+1}(x), \ldots, f_{n+p}(x)\}.$$

This function is in a class of index $\leqslant \alpha$. But then the function

$$\bar{f}_n(x) = \sup\{f_n(x), f_{n+1}(x), f_{n+2}(x), \ldots\} = \lim_{p\to\infty} \bar{f}_{n,p}(x)$$

is in a class with index $\leqslant \alpha + 1$ and, finally,

$$f(x) = \lim_{n\to\infty} \bar{f}_n(x)$$

is in a class with index $\leqslant \alpha + 2$.

We note one consequence of this theorem which is not intimately related to the theory under discussion, but is not devoid of interest.

THEOREM (G. VITALI). *Every measurable almost everywhere finite function $f(x)$, defined on $[a, b]$, is equivalent to some function $g(x)$ of Baire class not greater than two.*

In fact, according to Fréchet's theorem, there exists a sequence of continuous functions $\{\varphi_n(x)\}$ such that

$$\lim_{n\to\infty} \varphi_n(x) = f(x)$$

almost everywhere on $[a, b]$.

Setting

$$g(x) = \overline{\lim}\, \varphi_n(x),$$

we obtain the required function.

Let us now return to the main theme.

LEMMA 3. *If $f(x)$ is continuous on $[a, b]$, then for every $\varepsilon > 0$, there exists a polynomial $P(x)$ with rational coefficients such that*

$$|f(x) - P(x)| < \varepsilon$$

for all x in $[a, b]$.

In fact, by the Weierstrass theorem, $f(x)$ can be approximated to within $\varepsilon/2$ by means of some polynomial $Q(x)$. We obtain the polynomial $P(x)$ required by replacing the coefficients in $Q(x)$ by sufficiently proximate rational numbers.

COROLLARY. *Every function $f(x)$ of the first class can be written in the form*

$$f(x) = \lim P_n(x),$$

where $P_n(x)$ is a polynomial with rational coefficients.

LEMMA 4. *If the function $\psi(x)$, defined and finite on the closed interval $[a, b]$, has a finite number of points of discontinuity, then $\psi(x)$ is of the first Baire class.*

Proof. Let the points of discontinuity of $\psi(x)$ be

$$c_1 < c_2 < \ldots < c_m \qquad (a < c_k < b).$$

We enclose each of the points c_k in the interval $\left(c_k - \frac{1}{n}, c_k + \frac{1}{n}\right)$ and assume n so large that these intervals are disjoint and are contained in $[a, b]$. The function $f_n(x)$ which is equal to $\psi(x)$ in the exterior of all the intervals $\left(c_k - \frac{1}{n}, c_k + \frac{1}{n}\right)$ and at the points c_k, and linear on the closed intervals $\left[c_k - \frac{1}{n}, c_k\right]$, $\left[c_k, c_k + \frac{1}{n}\right]$, is evidently continuous. At the same time,

$$f_n(x) \to \psi(x)$$

as $n \to \infty$.

We leave to the reader the slight modifications required in the proof when a or b are also points of discontinuity.

COROLLARY. *The function $\theta(x)$, equal to unity at one point of $[a,b]$ and to zero at the remaining points, is a function of the first class.*

LEMMA 5. *If each number x in $[0,1]$ is expanded in a decimal fraction*

$$x = 0.a_1 a_2 a_3 \ldots$$

(where an eventual 9 for period is avoided for numbers < 1), then $a_k = a_k(x)$ is a function of the first class.

In fact, since $a_k(x)$ is constant in each of the intervals $\left(\frac{n}{10^k}, \frac{n+1}{10^k}\right)$, it has only a finite number of points of discontinuity.

After the discussion of this somewhat heterogeneous material, we can now proceed to the fundamental theorem on the existence of functions in each class H_α ($\alpha < \Omega$). To this end, we shall slightly modify the Baire classification; to wit, we shall consider the set of all polynomials with rational coefficients to be the null class H_0^*. Then the first class H_1^* will contain all other continuous functions and functions in the first class H_1; all remaining Baire classes will remain unchanged. As a matter of fact, we simply transfer a part of the set of continuous functions from the null class to the first class. For this modified classification, the following theorem is true.

THEOREM 4 (H. LEBESGUE). *For every $\alpha > 0$, there exists a function of two variables of the first or second number class,*

$$F_\alpha(x,t) \qquad (0 \leqslant x \leqslant 1, 0 \leqslant t \leqslant 1)$$

such that 1) *$F_\alpha(x,t)$ is a Baire function*; 2) *every Baire function $f(x)$ of class $< \alpha$ is obtained from $F_\alpha(x,t)$ by fixing the value of t, i.e.*

$$f(x) = F_\alpha(x, t_0) \qquad (0 \leqslant x \leqslant 1).$$

Such a function $F_\alpha(x,t)$ is called the *universal* function for the set of Baire functions of classes $< \alpha$.

Proof. Let the function $\theta_n(t)$ equal 1 for $t = \frac{1}{n}$ and 0 for all other t in $[0,1]$. We enumerate all polynomials with rational coefficients $P_1(x), P_2(x), P_3(x), \ldots$ and set $F_1(x,t) = \sum_{n=1}^{\infty} P_n(x)\,\theta_n(t)$.

This is a Baire function. In fact, each of the factors $P_n(x)$ and $\theta_n(t)$ can be considered as a function of two variables and this function is a Baire function. But then $P_n(x)\,\theta_n(t)$ is also a (finite) Baire function, and all partial sums of the series will be Baire functions along with it. Since all terms of the series, except perhaps one, are zeros, the series converges at every point (x,t) of the square $0 \leqslant x, t \leqslant 1$, and its sum is a Baire function.

Any function $f(x)$ of the null class H_0^* can be written in the form

$$f(x) = F_1(x, t_0).$$

In fact, if $f(x)$ is a polynomial $P_k(x)$, then

$$F_1\left(x, \frac{1}{k}\right) = P_k(x) = f(x).$$

Thus, our theorem is established for $\alpha = 1$. Now we assume that the functions $F_\beta(x,t)$ have been constructed for all $\beta < \alpha$; we shall construct $F_\alpha(x,t)$. To this end, we distinguish two cases:

1) α is a number of the first type. Then $\alpha = \beta + 1$, and the function $F_\beta(x, t)$ exists. We define the following sequence of functions in $t (0 \leqslant t \leqslant 1)$: expand t in a decimal fraction (without 9 in period for $t < 1$)

$$t = 0.a_1 a_2 a_3 \ldots$$

and set

$$h_1(t) = 0.a_1 a_3 a_5 \ldots$$
$$h_2(t) = 0.a_2 a_6 a_{10} \ldots$$
$$h_3(t) = 0.a_4 a_{12} a_{20} \ldots$$
$$\ldots\ldots\ldots$$

Since $h_n(t)$ is the sum of a uniformly convergent series

$$h_n(t) = \sum_{k=1}^{\infty} \frac{a_{2^{n-1}(2k-1)}(t)}{10^k},$$

the terms of which are functions of the first class, $h_n(t)$ is a function of the first class.

Let

$$F_\alpha(x, t) = \overline{\lim_{k \to \infty}} F_\beta[x, h_k(t)].$$

It is clear that $F_\alpha(x, t)$ is a Baire function; we shall show that it is the universal function for the set of functions of classes $< \alpha$.

In fact, if $f(x)$ is a function of class $< \alpha$, then

$$f(x) = \lim_{k \to \infty} f_k(x),$$

where each of the functions $f_k(x)$ is in a class $< \beta$ and, hence, representable in the form

$$f_k(x) = F_\beta(x, t_k) \qquad (0 \leqslant t_k \leqslant 1).$$

We can easily find a t^* in $[0, 1]$ such that for every k,

$$h_k(t^*) = t_k \qquad (k = 1, 2, 3, \ldots).$$

But then

$$F_\alpha(x, t^*) = \overline{\lim_{k \to \infty}} F_\beta(x, t_k) = \overline{\lim_{k \to \infty}} f_k(x) = f(x),$$

and, consequently, $F_\alpha(x, t)$ satisfies all conditions of the theorem.

2) Now let α be a number of the second type. Then there exists a sequence $\beta_1 < \beta_2 < \beta_3 < \ldots$ such that α is the smallest number following all numbers of this sequence.

We set

$$F_\alpha(x, t) = \overline{\lim_{k \to \infty}} F_{\beta_k}[x, h_k(t)],$$

where the $h_k(t)$ are the same functions as above. The function $F_\alpha(x, t)$ is a Baire function.

Let $f(x)$ be a function of class $< \alpha$. Then $f(x)$ is in a class H_γ, where $\gamma < \alpha$. For $k > k_0$, we have $\beta_k > \gamma$ and, hence, $f(x)$ can be written in the form

$$f(x) = F_{\beta_k}(x, t_k) \qquad (k > k_0).$$

Taking a t^* in [0, 1] such that

$$h_k(t^*) = t_k \qquad (k > k_0)$$

(obviously, this can be easily done so that $h_1(t^*), \ldots, h_{k_0}(t^*)$ can be taken arbitrarily), we have

$$F_\alpha(x, t^*) = \overline{\lim} F_{\beta_k}(x, t_k) = \overline{\lim} f(x) = f(x),$$

from which it is clear that $F_\alpha(x, t)$ is the required function. This completes the proof of the theorem.

THEOREM 5. *None of the classes H_α is void.*

Proof. Assume some class H_α is void. Then all the following classes are void, a fortiori, and all Baire functions belong to classes with indices $< \alpha$.

We shall construct the function $F_\alpha(x, t)$ which was discussed in connection with the proof of the preceding theorem.

If we set

$$\Phi(x, t) = [F_\alpha(x, t)]_0^1,$$

then every Baire function $f(x)$, whose values lie between 0 and 1, can be written in the form $f(x) = \Phi(x, t_0)$ for some fixed t_0. Having noted this, we set

$$\varphi(x, t) = \lim_{n \to \infty} \frac{n\Phi(x, t)}{1 + n\Phi(x, t)}.$$

This function assumes only two values: 0 and 1. It is a Baire function and every Baire function, which assumes only the values 0 and 1, can be written in the form $\varphi(x, t_0)$.

In particular, the function $1 - \varphi(x, x)$ can be written in this form. But this leads at once to an absurdity because the equality

$$1 - \varphi(x, x) = \varphi(x, t_0)$$

becomes

$$\varphi(t_0, t_0,) = \tfrac{1}{2}$$

for $x = t_0$. This completes the proof of the theorem.

Returning to Theorem 4, we mention the interesting work of L. V. Kantorovič,[5] where the following result, among others, is established: for the set of Baire functions of classes $< \alpha$, there exists a universal function $F_\alpha(x, t)$ which itself, being a Baire function, is in the class H_α; but such a universal function for the set of functions of classes $\leqslant \alpha$ does not exist.

§3. FUNCTIONS OF THE FIRST CLASS

In this section, we shall especially stop to consider some properties of functions of the first class.

LEMMA 1. 1) *A closed set is a set of type G_δ*; 2) *an open set is a set of type F_σ*; 3) *the difference of two closed sets is a set of type F_σ.*

Proof. Let F be a closed set. Denote the distance of the point x to F by $\rho(x, F)$ and set

$$G_n = Z\left(\rho(x, F) < \frac{1}{n}\right).$$

We saw (Chapt. II, §4, Lemma 1) that G_n is an open set. But it is easy to see that

$$F = \prod_{n=1}^{\infty} G_n,$$

and so F is a set of type G_δ.

---

[5] L. V. Kantorovič, *On universal functions*, Journal of the Leningrad Phys.-Math. Society, Vol. II, No. 2, 1929, pp. 13-21.

Proceeding to the proof of 2), we consider any open set G. Then its complement CG is closed and, according to what has been proved above,

$$CG = \prod_{n=1}^{\infty} G_n,$$

where the G_n are open. It follows that

$$G = \sum_{n=1}^{\infty} CG_n,$$

where CG_n is the closed complement of the set G_n. This means that G is of type F_σ.

Finally, let F_1 and F_2 be closed sets. Then the set $F_1 - F_2$ can be written in the form $F_1 - F_2 = F_1 \cdot CF_2$ which, being the intersection of a closed set and a set of type F_σ, is itself an F_σ.

LEMMA 2. *If a set M admits of the representation*

$$M = A_1 + A_2 + \ldots + A_n,$$

where the A_k are sets of type F_σ, then there exists another representation

$$M = B_1 + B_2 + \ldots + B_n,$$

in which the B_k are also sets of type F_σ, $B_k \subset A_k (k = 1, \ldots, n)$ and, furthermore, the sets B_k are pairwise disjoint.

Proof. It is clear that

$$M = \sum_{k=1}^{\infty} F_k,$$

where the F_k are closed sets and each of the sets F_k is contained entirely in some one of the sets A_i. We set

$$S_1 = F_1, \quad S_k = F_k - (F_1 + \ldots + F_{k-1}).$$

The sets S_k are F_σ sets, they are pairwise disjoint, and

$$M = \sum_{k=1}^{\infty} S_k.$$

We decompose the set $T = \{S_k\}$ into n parts $T_1, T_2, \ldots, T_n$ putting those sets S_k into T_1 which are contained in A_1, those $S_k \in T - T_1$ which are contained in A_2 into T_2, and so on. Setting

$$B_i = \sum_{S_k \in T_i} S_k,$$

we obtain the required decomposition of the set M.

LEMMA 3. *Let the function $f(x)$ be defined on the segment $E = [a, b]$ and let $f(x)$ assume a finite number of finite values $c_1 < c_2 < \ldots < c_n$.*

If each of the sets $E(f = c_k)$ is a set of type F_σ, then $f(x)$ is a function of class not greater than the first.

Proof. Let

$$E(f = c_k) = \sum_{i=1}^{\infty} F_i^{(k)},$$

where the $F_i^{(k)}$ are closed sets. We set

$$\Phi_m^{(k)} = \sum_{i=1}^{m} F_i^{(k)}, \quad \Phi_m = \sum_{k=1}^{n} \Phi_m^{(k)}$$

and introduce the function $\varphi_m(x)$ by setting

$$\varphi_m(x) = c_k \qquad (x \in \Phi_m^{(k)}, k = 1, 2, \ldots, n).$$

The function $\varphi_m(x)$ is defined on the closed set Φ_m and, by Lemma 1, §4, Chapt. IV, it is a continuous function. Using Lemma 2 of the same section, it is possible to construct a continuous function $\psi_m(x)$ on $[a, b]$ which coincides with $\varphi_m(x)$ on the set Φ_m. It is not difficult to verify that

$$\lim_{m \to \infty} \psi_m(x) = f(x)$$

at every point of $[a, b]$, from which the lemma follows.

THEOREM 1 (H. LEBESGUE). *A necessary and sufficient condition that a function $f(x)$, defined on the segment $E = [a, b]$, be a function of class not greater than the first, is that the sets $E(f > A)$, $E(f < A)$ be sets of type F_σ for arbitrary A.*

Proof. If $f(x)$ is a function of class not greater than the first, then

$$f(x) = \lim_{n \to \infty} f_n(x),$$

where the $f_n(x)$ are continuous. We set

$$F_n^{(k)} = E\left(f_n \leqslant A - \frac{1}{k}\right), \qquad S_m^{(k)} = \prod_{n=m}^{\infty} F_n^{(k)}.$$

In the proof of Theorem 8, §1, Chapt. IV, it was established that if the fun ction $f(x)$ is continuous, then the set $E(f \leqslant A)$ is closed. This means that the sets $F_n^{(k)}$, and also the sets $S_m^{(k)}$, are closed. Noting that

$$E(f < A) = \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} S_m^{(k)},$$

we can easily verify that $E(f < A)$ is a set of type F_σ. For $E(f > A)$, the reasoning is analogous.

Proceeding to the proof of the sufficiency of the condition of the theorem, we first assume that $f(x)$ is a bounded function: $l < f(x) < L$.

We subdivide the segment $[l, L]$ into n equal parts by means of the points

$$c_0 = l < c_1 < c_2 < \ldots < c_n = L \qquad \left(c_{k+1} - c_k = \frac{L - l}{n}\right)$$

and set

$$A_k = E(c_{k-1} < f < c_{k+1}) \qquad (k = 1, 2, \ldots, n-1)$$
$$A_0 = E(f < c_1), \qquad A_n = E(f > c_{n-1}).$$

All these sets are F_σ sets and

$$E = A_0 + A_1 + \ldots + A_n.$$

On the basis of Lemma 2, there exists another decomposition,

$$E = B_0 + B_1 + \ldots + B_n,$$

in which the sets B_k are also F_σ sets, *pairwise disjoint*, and such that $B_k \subset A_k$.

We introduce the function $f_n(x)$ by setting

$$f_n(x) = c_k \text{ for } x \in B_k \quad (k = 0, 1, \ldots, n).$$

According to Lemma 3, $f_n(x)$ is of class not greater than the first. We choose an arbitrary point $x_0 \in E$. Then

$$x_0 \in B_k \subset A_k.$$

This means that

$$f_n(x_0) = c_k, \quad c_{k-1} < f(x_0) < c_{k+1}.$$

It is now clear that

$$|f_n(x_0) - f(x_0)| < \frac{L-l}{n}.$$

Thus, as $n \to \infty$, the functions $f_n(x)$ converge to the function $f(x)$ *uniformly* and hence the latter is a function of class not greater than the first.

Going over to the general case, we set

$$g(x) = \arctan f(x).$$

This function is bounded. For $-\frac{\pi}{2} \leqslant A < \frac{\pi}{2}$,

$$E(g > A) = E(f > \tan A);$$

and if $A \geqslant \frac{\pi}{2}$, the set $E(g > A)$ is void. Finally, if $A < -\frac{\pi}{2}$, then $E(g > A) = [a, b]$. Thus, the set $E(g > A)$ is an F_σ set for all A. The same is true for $E(g < A)$. Hence, $g(x)$ is a function of class not greater than the first, and, therefore,

$$f(x) = \tan [g(x)]$$

is a function of class not greater than the first.[6]

R. Baire discovered another, very interesting, characteristic of functions of the first class. We need some new concepts and facts to discuss his theorems.

DEFINITIONS. Let A and B be two point sets such that $A \subset B$. 1) If every interval, containing at least one point of B, contains points of B which are not in the closure $\bar{A}$ of the set A, then we say that A is *nowhere dense* on the *set* B. 2) If A is representable as the sum of a denumerable number of sets, each of which is nowhere dense on the set B, then we say that A is a *set of the first category* on the set B.

THEOREM 2. *A non-void closed set F is not a set of the first category on itself.*

Proof. We assume, on the contrary, that F can be written in the form

$$F = A_1 + A_2 + A_3 + \ldots,$$

where each of the sets A_k is nowhere dense on the set F. Then there exists a point $x_1 \in F$ which is not a point in the closure $\bar{A}_1$ of the set A_1. Therefore, a segment $[x_1 - \delta_1, x_1 + \delta_1]$ can be found not containing any point of A_1, where we can assume that $\delta_1 < 1$.

In the interval $(x_1 - \delta_1, x_1 + \delta_1)$ there exists a point $x_2 \in F_2$ not in the set $\bar{A}_2$. This means that a segment $[x_2 - \delta_2, x_2 + \delta_2]$ can be found which does not contain any point of A_2. We can assume that $\delta_2 < \frac{1}{2}$ and that

$$[x_2 - \delta_2, x_2 + \delta_2] \subset [x_1 - \delta_1, x_1 + \delta_1].$$

---

[6] The function $g(x)$ can be written in the form

$$g(x) = \lim_{n \to \infty} g_n(x),$$

where the $g_n(x)$ are continuous functions which satisfy the condition

$$-\frac{\pi}{2} + \frac{1}{n} \leqslant g_n(x) \leqslant \frac{\pi}{2} - \frac{1}{n}.$$

Then

$$f(x) = \lim_{n \to \infty} \tan [g_n(x)],$$

where $\tan [g_n(x)]$ is a continuous function.

Continuing this process, we construct a sequence of points

$$x_1, x_2, \ldots, x_n, \ldots$$

in F, and a sequence of nested segments

$$[x_1 - \delta_1, x_1 + \delta_1] \supset [x_2 - \delta_2, x_2 + \delta_2] \supset \ldots \supset [x_n - \delta_n, x_n + \delta_n] \supset \ldots$$

such that the segment $[x_n - \delta_n, x_n + \delta_n]$ does not contain any point of the set A_n and $\delta_n < \frac{1}{n}$.

Let x_0 be the point common to all the segments $[x_n - \delta_n, x_n + \delta_n]$.

Obviously,

$$x_0 = \lim x_n,$$

and hence $x_0 \in F$. At the same time, x_0 is not in any of the sets A_n. This contradiction proves the theorem.

COROLLARY. *If a non-void closed set F is the sum of a denumerable number of closed sets, $F = F_1 + F_2 + F_3 + \ldots$, then there exist an interval (λ, μ) containing points of the set F and an n such that $(\lambda, \mu) F \subset F_n$.*

In fact, at least one of the sets in $F_1 + F_2 + F_3 + \ldots$, say F_n, will not be nowhere dense on the set F. But this means that among the intervals containing points of F, an interval (λ, μ) can be found such that all points of F in it are in F_n, because the latter set, being closed, coincides with its closure.

Let a function $f(x)$ be defined on some set A and let B be a subset of A. We consider the function $f(x \mid B)$ defined only at points of the set B and coinciding with the function $f(x)$ at these points. We shall say that the function $f(x \mid B)$ is *induced* by the function $f(x)$ in the set B. It is easy to see that if the initial function $f(x)$ is continuous on the set A, then the induced function $f(x \mid B)$ is continuous on the set B.

THEOREM 3 (R. BAIRE). *Let a finite[7] function $f(x)$ of the first class be defined on the segment $[a, b]$. Every closed set $F \neq 0$, contained in $[a, b]$, contains points of continuity of the induced function $f(x \mid F)$.*

Proof. If the set F has at least one isolated point, this point will be the point of continuity sought. Setting this trivial case aside, we shall assume that F is a perfect set.

Let D be a segment contained in $[a, b]$ and suppose there is at least one point (and, hence, an infinite set of points) of the set F in the interior of D.

We shall show that there exists a segment d lying in the interior[8] of D, containing points of the set F in its interior and such that the oscillation of $f(x)$ on the set F is less than an arbitrary prescribed number.

To this end, we first of all assure ourselves of the existence of a segment $E \subset D$ such that the set EF is a (non-void) perfect set. Let the endpoints of D be A and B, so that $D = [A, B]$. The infinite set DF is closed and none of its points, except A and B, can be an isolated point. We shall assume that A is an isolated point of the set DF. Then the set $DF - \{A\}$, obtained from DF by removing the point A, is closed. Let the leftmost point of this set be A_1; we set $D_1 = [A_1, B]$. The only isolated point in the closed set D_1F can be B. If B is in fact such a point, we remove it from the set D_1F and denote the

---

[7] The finiteness condition on $f(x)$ is essential because a function which equals $+\infty$ everywhere is a function of the first class and hence has no points of continuity at all.

[8] We shall say that the segment $[\lambda, \mu]$ lies in the *interior* of the segment $[\sigma, \tau]$ if $\sigma < \lambda < \mu < \tau$.

rightmost point of the remaining (closed) set by B_1. The segment $E=[A_1, B_1]$ will be such that EF is a perfect set.

By hypothesis, $f(x)$ can be written in the form

$$f(x)=\lim_{n\to\infty} f_n(x),$$

where the $f_n(x)$ are continuous functions. We take $\varepsilon>0$ and set

$$A_{n,m}=E(|f_n-f_{n+m}|\leqslant\varepsilon) \qquad (n=1,2,\ldots;m=1,2,\ldots).$$

Obviously the set $A_{n,m}$ is closed. Further, let

$$B_n=\prod_{m=1}^{\infty} A_{n,m}.$$

This is also a closed set. It is easy to see that

$$E=\sum_{n=1}^{\infty} B_n.$$

In fact, if $x_0\in E$, then the sequence $\{f_n(x_0)\}$ converges and therefore, for sufficiently large n and arbitrary m, we have

$$|f_n(x_0)-f_{n+m}(x_0)|<\varepsilon,$$

so that $x_0\in B_n$ and, hence, $E\subset\sum B_n$. The reverse inclusion is trivial.

But the equality just established implies that

$$EF=\sum_{n=1}^{\infty} FB_n.$$

By the preceding corollary, there exist an interval (λ,μ) containing points of the set EF and a natural number n such that $(\lambda,\mu)EF\subset FB_n$.

If $x\in(\lambda,\mu)EF$, then

$$|f_n(x)-f_{n+m}(x)|<\varepsilon$$

for arbitrary m.

Passing to the limit as $n\to\infty$ in the last inequality, we find that

$$|f_n(x)-f(x)|\leqslant\varepsilon.$$

The set EF is a perfect set; the interval (λ,μ) contains at least one point of this set. This means that the set $(\lambda,\mu)EF$ is infinite. Let x_0 be a point of the set $(\lambda,\mu)EF$, which is distinct from the endpoints of the segment E. We take a segment d, containing the point x_0 in its interior and so small that 1) it is contained in the interval (λ,μ), 2) it lies in the interior of the segment E (and, a fortiori, in the interior of the segment D), and 3) the oscillation of the (continuous) function $f_n(x)$ on the segment d is less than ε.

Let x_1 and x_2 be two points of the set Fd. Then

$$|f_n(x_1)-f(x_1)|\leqslant\varepsilon,\ |f_n(x_1)-f_n(x_2)|\leqslant\varepsilon,\quad |f_n(x_2)-f(x_2)|\leqslant\varepsilon,$$

from which it follows that

$$|f(x_1)-f(x_2)|<3\varepsilon,$$

i.e. the oscillation of $f(x)$ on the set Fd is less than 3ε.

We have thus established that for every segment $D\subset[a,b]$, containing points of F in its interior, there exists another segment d, lying in the interior of D, which also contains points of F in its interior and is such that the oscillation of $f(x)$ on Fd is as small as we please.

Having proved this, we now take a segment $d_1\subset[a,b]$, containing points of F in

its interior, of length $md_1 < 1$, and such that the oscillation of $f(x)$ on the set Fd_1 is less than 1.

Next, we find a segment d_2, lying in the interior of d_1, containing points of F in its interior, of length $md_2 < \frac{1}{2}$, and such that the oscillation of $f(x)$ on the set Fd_2 is less than $\frac{1}{2}$.

Continuing this process, we construct a sequence of segments

$$d_1 \supset d_2 \supset d_3 \supset \ldots \qquad \left(md_n < \frac{1}{n}\right),$$

each of which lies in the interior of the preceding, contains points of the set F in its interior, and is such that the oscillation of $f(x)$ on the set Fd_n is less than $\frac{1}{n}$.

Let ξ be the point common to all the segments d_n. It evidently belongs to the set F. It is easy to see that the induced function $f(x \mid F)$ is continuous at the point ξ. This completes the proof of the theorem.

It turns out that this theorem admits of a converse. To prove this, we need the

CANTOR-BAIRE STATIONARY PRINCIPLE. *Let a closed set F_α correspond to every ordinal number α of the first or second number class, where $\alpha < \beta$ implies $F_\alpha \supset F_\beta$:*

$$F_0 \supset F_1 \supset \ldots \supset F_\omega \supset \ldots \supset F_\alpha \supset \ldots \qquad (\alpha < \Omega).$$

Then all sets of the chain $\{F_\alpha\}$, beginning with some one of them, coincide, i.e. a number $\mu < \Omega$ can be found such that

$$F_\mu = F_{\mu+1} = F_{\mu+2} = \ldots. \qquad (*)$$

Proof.[9] We write the (denumerable!) set of all intervals with rational endpoints as a sequence $\delta_1, \delta_2, \delta_3, \ldots$. This enables us to assign to each set of points E on the real line a set $S(E)$ consisting of all natural numbers k which satisfy the relation $E\delta_k \neq 0$.

It is easy to see that $A \subset B$ implies $S(A) \subset S(B)$. Further, if the sets F and F^* are closed and $F \neq F^*$, then[10] $S(F) \neq S(F^*)$. In fact, suppose $x_0 \in F - F^*$. Then all sufficiently short intervals which contain x_0 will not intersect F^*. If δ_i is one of these intervals, then $i \in S(F) - S(F^*)$.

Having noted this, we set $S(F_\alpha) = S_\alpha$. This leads to the following chain of sets:

$$S_0 \supset S_1 \supset S_2 \supset \ldots \supset S_\alpha \supset \ldots \qquad (\alpha < \Omega).$$

Let K be the intersection of all the sets of this chain. If K consists of all the natural numbers then all the sets of the chain are also of the same sort. This means that all the sets F_α, beginning with F_0, coincide (each of them coincides with the entire real line). Setting aside this trivial case, we consider the set M of all natural numbers which do not occur in K. If $m \in M$, then numbers $\alpha < \Omega$ can be found such that $m \bar{\in} S_\alpha$. We denote the smallest of them by α_m. This gives rise to a finite or denumerable set of numbers $\{\alpha_m\}$. Since all of them are less than Ω, an ordinal number μ can be found which still belongs to the first or second number class and is larger than all the α_m. If $m \bar{\in} K$, then $m \bar{\in} S_{\alpha_m}$ and a fortiori $m \bar{\in} S_\mu$. It follows that $K = S_\mu$. Thus,

---

[9] This proof was communicated to the author by Academician P. S. Aleksandrov.

[10] The assertion is not true without the condition that both sets F and F^* are closed. For example, if R and Z are the sets of all rational numbers and of all real numbers, respectively, then $S(R) = S(Z) = \{1, 2, 3, \ldots\}$.

$$S_\mu = S_{\mu+1} = S_{\mu+2} = \ldots,$$

and this is equivalent to relation (*).

THEOREM 4 (R. BAIRE). *Let the function $f(x)$ be defined on the segment $[a,b]$. If there are points of continuity of the induced function $f(x \mid F)$ on every closed set F, $f(x)$ is continuous or of the first class.*

Proof. By virtue of Lebesgue's theorem, it will suffice to show that the sets

$$E(f > A), \quad E(f < A)$$

are sets of type F_σ for arbitrary A.

Let p and q be two numbers, $p < q$. We set

$$P = E(f > p), \quad Q = E(f < q).$$

Obviously,

$$E = P + Q.$$

Let F be a non-void closed set contained in E. We denote any point of continuity of the induced function $f(x \mid F)$ by x_0. It is clear that at least one of the inequalities

$$f(x_0) > p, \quad f(x_0) < q$$

is satisfied, for instance, $f(x_0) > p$. Then there exists an interval δ containing the point x_0 and so small that $f(x) > p$ for all points of the set $F\delta$. We set $F^* = F - F\delta$; this is a closed set because $F^* = F \cdot C\delta$. Here $F - F^* = F\delta \subset P$. If $f(x_0)$ were $< q$, we could find, in an analogous manner, a closed set $F^* \subset F$ of such a nature that $F - F^* \subset Q$. So, for any non-void closed set F, a closed subset F^* of F can be found such that the set $F - F^*$ is non-void and is contained entirely in P or Q.

Having noted this, we set $F_0 = [a,b]$ and find a closed set $F_1 \subset F_0$ of such a nature that $F_0 - F_1$ is not void and is contained entirely in P or Q. If F_1 is also not void, we find a closed subset F_2 of it such that $F_1 - F_2$ is non-void and is contained in P or Q. Continuing this process, we either arrive at a void set F_n or, for all natural numbers n, we construct sets F_n of such a nature that $F_n - F_{n+1}$ is not void and is contained in P or Q.

In the latter case, we set

$$F_\omega = \prod_{n=0}^{\infty} F_n.$$

If this set is still not void, the possibility arises of further constructing the sets $F_{\omega+1}, F_{\omega+2}, \ldots$. Let α be a number of the second class and assume that all sets F_β for $\beta < \alpha$ have already been defined and that all are non-void.

If α is a number of the first type, we shall denote by F_α a closed subset of the set $F_{\alpha-1}$ (where $\alpha - 1$ is the immediate predecessor of α) such that $F_{\alpha-1} - F_\alpha \neq 0$ and is contained entirely in P or Q. If, however, α is a number of the second type, then we set

$$F_\alpha = \prod_{\beta<\alpha} F_\beta.$$

Let us assume all the sets F_α, for $\alpha < \Omega$, are non-void. It is easy to see this assumption contradicts the Cantor-Baire Stationary Principle. In fact, according to this principle, a μ can be found such that

$$F_\mu = F_{\mu+1},$$

and, at the same time, the non-vacuousness of F_μ implies the non-vacuousness of the difference $F_\mu - F_{\mu+1}$.

Thus, the process of defining the sets F_α cannot be carried out for all numbers of the first two classes, and a $\lambda < \Omega$ necessarily exists such that

$$F_\alpha \neq 0 \quad (\alpha < \lambda), \quad F_\lambda = 0.$$

In this case, the initial segment $F_0 = [a, b]$ can be represented in the form

$$[a, b] = \sum_{\alpha < \lambda} [F_\alpha - F_{\alpha+1}].$$

In fact, if $x \in [a, b]$, then $\alpha \leqslant \lambda$ can be found such that $x \bar{\in} F_\alpha$ (for example, $\alpha = \lambda$). Let β be the first one of them. It is clear that β is a number of the first type (because if β were a number of the second type, then x, which is in all the F_α for $\alpha < \beta$, would also be in their intersection F_β). This means that $x \in F_{\beta-1} - F_\beta$ and

$$[a, b] \subset \sum_{\alpha < \lambda} [F_\alpha - F_{\alpha+1}].$$

The reverse inclusion is trivial. Each of the sets $F_\alpha - F_{\alpha+1}$ is in one of the sets P or Q. Denote the set of those $\alpha < \lambda$ for which

$$F_\alpha - F_{\alpha+1} \subset P$$

by T and let $U = W_\lambda - T$. It is clear that $UT = 0$ and

$$[a, b] = \sum_T [F_\alpha - F_{\alpha+1}] + \sum_U [F_\alpha - F_{\alpha+1}].$$

Each of the sets $F_\alpha - F_{\alpha+1}$ is a set of type F_σ; their sums

$$A = \sum_T [F_\alpha - F_{\alpha+1}], \quad B = \sum_U [F_\alpha - F_{\alpha+1}]$$

are also of the same type because each of the sets T or U is finite or denumerable. We note that $A \subset P, B \subset Q$ and the sets A and B are pairwise disjoint (because the sets $F_\alpha - F_{\alpha+1}$ and $F_\beta - F_{\beta+1}$ do not intersect for $\alpha \neq \beta$).

So, to every pair of numbers $p, q\ (p < q)$ there corresponds a decomposition of the segment $[a, b]$ into two disjoint sets of type F_σ,

$$[a, b] = A + B,$$

where

$$A \subset E(f > p), B \subset E(f < q).$$

Having established this fact, we now fix p and let q assume the sequence of values

$$q_1 > q_2 > q_3 > \ldots, \quad \lim q_n = p.$$

For every n, we have

$$[a, b] = A_n + B_n \qquad (A_n B_n = 0),$$

where A_n and B_n are sets of type F_σ and

$$A_n \subset E(f > p), \quad B_n \subset E(f < q_n).$$

Set

$$R = \sum_{n=1}^{\infty} A_n, \qquad S = \prod_{n=1}^{\infty} B_n.$$

Obviously $RS = 0$ and $[a, b] = R + S$.

The set R is an F_σ. We shall show that

$$E(f > p) = R.$$

In fact, if $f(x_0) > p$, then for sufficiently large n we have $f(x_0) > q_n$ and $x_0 \bar{\in} B_n$. This means that $x_0 \bar{\in} S$ and $x_0 \in R$.

It follows that

$$E(f>p) \subset R.$$

The reverse inclusion is obvious. So, $E(f>p)$ is an F_σ. The same is true of the sets $E(f<q)$. This completes the proof of the theorem.

We shall illustrate the theorems of Lebesgue and Baire by giving some examples.

I. *The characteristic function of a bounded closed set F is a function of the first class.*

Let $F \subset [a,b] = E$; let $f(x)$ equal 1 at points of F and 0 at points of the set $E - F$. Then

$$E(f>A) = \begin{cases} E, \text{ if } A<0 \\ F, \text{ if } 0 \leqslant A<1 \\ 0, \text{ if } A>1 \end{cases} \quad E(f<A) = \begin{cases} E, \text{ if } A>1 \\ E-F, \text{ if } 0<A \leqslant 1 \\ 0, \text{ if } A \leqslant 0. \end{cases}$$

In all cases the sets $E(f>A)$ and $E(f<A)$ are F_σ's. In order to make this clear, it suffices to refer to Lemma 1.

We note that the assertion can easily be proved without the aid of the preceding theory. If we denote the distance from the point x to the set F by $r(x)$, then the function $r(x)$ is continuous. At the same time,

$$f(x) = \lim_{n\to\infty} \frac{1}{1+nr(x)}.$$

II. *If $f(x)$ is a function having a denumerable set of points of discontinuity, then $f(x)$ is a function of the first class.*

In fact, if F is a closed set having isolated points, these points will also be points of continuity of the induced function $f(x \mid F)$. But if F is a perfect set, it is non-denumerable and, therefore, it contains points of continuity of the original function which are, a fortiori, points of continuity of the induced function.

In particular:

III. *Monotonic functions and functions of finite variation are functions of class not greater than the first.*

The following example is rather instructive.

IV. Let P_0 be the Cantor perfect set. Define two functions $f(x)$ and $g(x)$ on the segment $[0,1]$ by setting $f(x)$ equal to 1 at points of the set P_0 and 0 at points of its open complement $G_0 = [0,1] - P_0$; the function $g(x)$ is set equal to 1 at those, and only those, points of the set P_0 which are not endpoints of the complementary intervals, and $g(x) = 0$ at the remaining points of $[0,1]$.

It is easy to see that both $f(x)$ and $g(x)$ are continuous functions at points of the set G_0 and discontinuous at points of the set P_0. In other words, these functions have the same set of points of discontinuity. At the same time $f(x)$ *is a function of the first class* (being the characteristic function of the closed set P_0) but $g(x)$ is not of this class because the induced function $g(x \mid P_0)$ is discontinuous at every point of the set P_0.

V. *Dirichlet's function is a function of the second class.*

The Dirichlet function equals 1 at rational points of the segment $[0,1]$ and 0 at irrational points of $[0,1]$. It has absolutely no points of continuity and, therefore, it cannot be a function of the first class. At the same time, if we enumerate the rational points of the segment $[0,1]$: $r_1, r_2, r_3, \ldots$ and set

$$\varphi_n(x) = \begin{cases} 1 \text{ for } x = r_k & (k = 1, 2, \ldots, n) \\ 0 \text{ for remaining points of } [0, 1], \end{cases}$$

$\varphi_n(x)$, having a finite number of points of discontinuity, will be a function of the first class. Since the Dirichlet function is the limit of $\varphi_n(x)$ as $n \to \infty$, the assertion is proved.[11]

We note that the Dirichlet function is the characteristic function of the set of rational numbers which (being denumerable) is obviously an F_σ. This means that the characteristic function of a set of type F_σ is not always a function of the first class.

§4. SEMI-CONTINUOUS FUNCTIONS

We now stop to consider a special set of functions of the first class—these are the so-called semi-continuous functions. To this end, we must first extend to functions the concepts of limit superior and limit inferior which were established in §2 for sequences. For the sake of simplicity, we shall restrict the discussion to functions of one variable.

Let the function $f(x)$ be defined on the set E which has a limit point x_0. We take a number $\delta > 0$ and set

$$M_\delta(x_0) = \sup\{f(x)\}, \quad m_\delta(x_0) = \inf\{f(x)\} \quad [x \in (x_0 - \delta, \ x_0 + \delta) E].$$

The quantity $M_\delta(x_0)$ does not increase and $m_\delta(x_0)$ does not decrease with decreasing δ and, therefore, the (finite or infinite) limits

$$M(x_0) = \lim_{\delta \to 0} M_\delta(x_0), \ m(x_0) = \inf_{\delta \to 0} m_\delta(x_0)$$

exist; they are called[12] the *limit superior* and *limit inferior*, respectively, of the function $f(x)$ at the point x_0. These limits are denoted by

$$M(x_0) = \overline{\lim_{x \to x_0}} f(x), \quad m(x_0) = \underline{\lim_{x \to x_0}} f(x).$$

We note that $f(x)$ might not be defined at the point x_0 itself. But if $x_0 \in E$, then obviously

$$m(x_0) \leqslant f(x_0) \leqslant M(x_0).$$

THEOREM 1. *The limit superior $M(x_0)$ of the function $f(x)$ at the point x_0 is the greatest of the numbers which are limits of sequences of the form*

$$f(x_1), f(x_2), f(x_3), \ldots,$$

where $x_n \in E$, $x_n \to x_0$.

Proof. If $M(x_0) = -\infty$, then it is easy to verify that $f(x_n) \to M(x_0)$ for *every* sequence $\{x_n\} \subset E$ which converges to x_0. Let $-\infty < M(x_0) < +\infty$. Then an N can be found such that $M_{\frac{1}{n}}(x_0) < +\infty$ for $n > N$. For every natural $n > N$, a point $x_n \in E$ can be found in the interval $\left(x_0 - \frac{1}{n}, x_0 + \frac{1}{n}\right)$ such that

$$M_{\frac{1}{n}}(x_0) - \frac{1}{n} < f(x_n) \leqslant M_{\frac{1}{n}}(x_0).$$

---

[11] It is also not difficult to verify that the Dirichlet function $\psi(x)$ can be written in the form

$$\psi(x) = \lim_{m \to \infty} \{ \lim_{n \to \infty} [\cos(m! \pi x)]^{2n} \},$$

from which it is again clear that $\psi(x)$ is a function of the second class.

[12] The reader will recognize here the upper and lower Baire functions which were discussed in §4, Chapt. V.

It is clear that $f(x_n) \to M(x_0)$. On the other hand, if $B > M(x_0)$, then a $\delta > 0$ can be found such that $M_\delta(x_0) < B$. For all $x \in (x_0 - \delta, x_0 + \delta)$, we have $f(x) \leqslant M_\delta(x_0)$ and, therefore, no sequence $\{x_n\} \subset E$ can be found such that

$$x_n \to x_0, \qquad f(x_n) \to B.$$

Finally, if $M(x_0) = +\infty$, then $M_{\frac{1}{n}}(x_0) = +\infty$ for all natural n and, hence, a point $x_n \in E$ can be found in the interval $\left(x_0 - \frac{1}{n}, x_0 + \frac{1}{n}\right)$ for which $f(x_n) > n$. Then $f(x_n) \to M(x_0)$ and, in general, *larger* numbers do not exist. This completes the proof of the theorem. It is perfectly evident that an analogous theorem is valid for $m(x_0)$.

DEFINITION 1. A function $f(x)$, defined on the segment $[a, b]$, is said to be *lower semi-continuous* at the point x_0 of this segment if

$$\underline{\lim}_{x \to x_0} f(x) = f(x_0).$$

If

$$\overline{\lim}_{x \to x_0} f(x) = f(x_0),$$

then $f(x)$ is said to be *upper semi-continuous* at the point x_0.

We do not assume in this definition that the function $f(x)$ is finite either at the point x_0 itself or at the other points of the segment $[a, b]$. In particular, $f(x)$ is lower semi-continuous at every point x_0 for which $f(x_0) = -\infty$.

If $f(x)$ is continuous at the point x_0, then it will be both upper and lower semi-continuous at this point. Conversely, it follows from the upper and lower semi-continuity of $f(x)$ at the point x_0 [under the condition $f(x)$ is finite] that $f(x)$ is continuous at the point x_0. These assertions are only another formulation of Theorem 1 of §4, Chapt. V.

In the sequel, we shall consider mainly functions which are lower semi-continuous. Everything discussed can easily be extended to upper semi-continuous functions if we note that the lower semi-continuity of a function $f(x)$ at the point x_0 is equivalent to the upper semi-continuity of the function $-f(x)$ at the same point x_0.

The concept of semi-continuity of a function can also be defined in a somewhat different form:

THEOREM 2. *Suppose the function $f(x)$ is defined on $[a, b]$ and let $x_0 \in [a, b]$. A necessary and sufficient condition that $f(x)$ be lower semi-continuous at x_0 is that for an arbitrary sequence of points $x_n \in [a, b]$, which converges to x_0, we have*

$$f(x_0) \leqslant \underline{\lim}_{n \to \infty} f(x_n). \tag{1}$$

In fact, suppose $f(x)$ is lower semi-continuous at x_0 and let $x_n \to x_0$ $(x_n \in [a, b])$. Then $\{x_n\}$ contains a subsequence $\{x_{n_k}\}$ such that $f(x_{n_k})$ tends to $\underline{\lim}_{n \to \infty} f(x_n)$. By Theorem 1, this implies that

$$f(x_0) = \underline{\lim}_{x \to x_0} f(x) \leqslant \lim_{k \to \infty} f(x_{n_k}) = \underline{\lim}_{n \to \infty} f(x_n).$$

Conversely, let (1) be satisfied for every sequence $\{x_n\} \subset [a, b]$ which converges to x_0. Taking, in particular, that sequence for which

$$f(x_n) \to \underline{\lim}_{x \to x_0} f(x),$$

we obtain that $f(x_0) \leqslant \varlimsup_{x \to x_0} f(x)$ and since it is known that $\varlimsup_{x \to x_0} f(x) \leqslant f(x_0)$, $f(x)$ is lower semi-continuous at x_0.

THEOREM 3. *Let $f(x)$ be defined on $[a, b]$, $x_0 \in [a, b]$ and $f(x_0) > -\infty$. A necessary and sufficient condition that $f(x)$ be lower semi-continuous at x_0 is that to every $A < f(x_0)$ there correspond a $\delta > 0$ such that $f(x) > A$ provided $|x - x_0| < \delta$ (and $x \in [a, b]$).*

In fact, let $f(x)$ be lower semi-continuous at x_0 and let $A < f(x_0)$. Since

$$f(x_0) = \varliminf_{x \to x_0} f(x) = m(x_0) = \lim_{\delta \to 0} m_\delta(x_0),$$

a $\delta > 0$ can be found such that $m_\delta(x_0) > A$. Since $f(x) \geqslant m_\delta(x_0)$ for $|x - x_0| < \delta$ (and $x \in [a, b]$), the δ found is the one required. Conversely, let $f(x)$ possess the property stated in the theorem. Taking an arbitrary $A < f(x_0)$ and finding a δ corresponding to this A, we shall obviously have $m_\delta(x_0) \geqslant A$, from which it follows, a fortiori, that $m(x_0) \geqslant A$.

Increasing A and taking the limit as $A \to f(x_0)$, we find that

$$m(x_0) \geqslant f(x_0);$$

in combination with the fact that $m(x_0) \leqslant f(x_0)$, this implies that $f(x)$ is lower semi-continuous at x_0.

If we assume that $f(x_0)$ is a *finite* number, then the theorem just proved can be stated in the following form:

THEOREM 4. *A necessary and sufficient condition that $f(x)$ be lower semi-continuous at the point x_0 is that to every $\varepsilon > 0$, there correspond a $\delta > 0$ such that $f(x_0) - \varepsilon < f(x)$ provided $|x - x_0| < \delta$ (and $x \in [a, b]$).*

Here the connection between the concepts of semi-continuity and continuity is particularly clear.

THEOREM 5. *Suppose $f(x)$ and $g(x)$ are defined on the segment $[a, b]$ and lower semi-continuous at the point x_0. If the sum $f(x) + g(x)$ is defined[13] for all $x \in [a, b]$, then it is lower semi-continuous at x_0.*

Proof. We may assume $f(x_0) + g(x_0) > -\infty$ because otherwise there is nothing to prove. But then each of the numbers $f(x_0)$ and $g(x_0)$ separately is also distinct from $-\infty$. We take any number $A < f(x_0) + g(x_0)$. Then there exist B and C such that

$$B < f(x_0), C < g(x_0), B + C > A.$$

By Theorem 3, there exists a $\delta > 0$ such that $f(x) > B, g(x) > C$ for $|x - x_0| < \delta$ (and $x \in [a, b]$). But then, a fortiori, for these x, $f(x) + g(x) > A$, from which again by Theorem 3 it follows the sum $f(x) + g(x)$ is lower semi-continuous at the point x_0.

THEOREM 6. *Let an increasing sequence of functions*

$$f_1(x) \leqslant f_2(x) \leqslant f_3(x) \leqslant \ldots$$

be defined on $[a, b]$; assume also that all the functions $f_n(x)$ are lower semi-continuous at the point $x_0 \in [a, b]$. Then the function

---

[13] This means that $f(x)$ and $g(x)$ do not assume infinite values of opposite sign at the same point. If, for example, $f(x) = -\infty$ for $0 \leqslant x \leqslant 1$ and

$$g(x) = \begin{cases} -\infty & \text{for } x = 0 \\ +\infty & \text{for } 0 < x \leqslant 1, \end{cases}$$

then both these functions are lower semi-continuous for $x = 0$ but it is impossible to talk about their sum for $x > 0$.

$$f(x) = \lim_{n\to\infty} f_n(x)$$

is also lower semi-continuous at x_0.

Proof. We may assume that $f(x_0) > -\infty$. If $A < f(x_0)$, then $f_n(x_0) > A$ for sufficiently large n. Keeping such an n fixed, we find a $\delta > 0$ for which $f_n(x) > A$ provided $|x - x_0| < \delta$ $(x \in [a, b])$. Since $f(x) \geqslant f_n(x)$, for these x we have a fortiori that $f(x) > A$, from which the theorem follows.

COROLLARY. *If all the terms of the series*

$$u_1(x) + u_2(x) + u_3(x) + \dots$$

are non-negative and lower semi-continuous at the point x_0, *then the sum*[14] *of this series is also lower semi-continuous at* x_0.

Till now we considered functions which were semi-continuous *at a point.* Now we go over to functions which are semi-continuous *on a segment.*

DEFINITION 2. A function $f(x)$ is said to be *lower semi-continuous on the segment* $[a, b]$ if it is defined and lower semi-continuous at each point of this segment. Upper semi-continuity on a segment is defined analogously.

THEOREM 7. *A necessary and sufficient condition that a function* $f(x)$, *defined on the segment* $E = [a, b]$, *be lower semi-continuous on* E *is that the set*

$$F(c) = E(f \leqslant c)$$

be closed[15] *for arbitrary real* c.

Proof. Let $f(x)$ be lower semi-continuous on $[a, b]$. If $x_n \in F(c)$ and $x_n \to x_0$, then by Theorem 2

$$f(x_0) \leqslant \varliminf_{n\to\infty} f(x_n),$$

from which it follows that $x_0 \in F(c)$; hence $F(c)$ is a closed set.

Now we shall assume that all the sets $F(c)$ are closed and verify that the function $f(x)$ is lower semi-continuous at a point x_0 taken at random on the segment $[a, b]$. To this end, we set

$$A = \varliminf_{x\to x_0} f(x).$$

By Theorem 1, there exists a sequence of points $x_n \in [a, b]$ such that

$$x_n \to x_0, \quad f(x_n) \to A.$$

Let us assume that

$$f(x_0) > A. \tag{2}$$

If we denote any number lying between A and $f(x_0)$ by c, $A < c < f(x_0)$, then $f(x_n) < c$ for all sufficiently large n, i.e. $x_n \in F(c)$. But then $x_0 \in F(c)$ also, i.e. $f(x_0) \leqslant c$, which contradicts the very choice of the number c. This means (2) is impossible and hence

$$f(x_0) \leqslant \varliminf_{x\to x_0} f(x).$$

This completes the proof of the theorem.

Example. Let the set S be the intersection of the segment $[a, b]$ with an *open* set G. If $\varphi(x)$ is the characteristic function[16] of the set S, then $\varphi(x)$ is lower semi-continuous on $[a, b]$.

---

[14] The convergence of the series is not assumed. Its sum equals $+\infty$ at points of divergence.

[15] This implies, among other things, that such a function is *measurable.*

[16] Defined on $[a, b]$, i.e. $\varphi(x) = 1$ for $x \in S$ and $\varphi(x) = 0$ for $x \in [a, b] - S$.

In fact, in this case

$$F(c)=\begin{cases}[a,b] & \text{if } c\geqslant 1,\\ [a,b]-G & \text{if } 0\leqslant c<1,\\ 0 & \text{if } c<0.\end{cases}$$

In all these cases, $F(c)$ is closed.

THEOREM 8. *A function which is lower semi-continuous on a segment attains its minimum value on this segment.*[17]

Proof. Suppose $f(x)$ is lower semi-continuous on the segment $[a,b]$ and that $m=\inf\{f(x)\}$. Then a sequence $\{x_n\}$ can be selected from $[a,b]$ for which

$$\lim_{n\to\infty} f(x_n)=m.$$

Going over to a subsequence if necessary, it is possible to have the sequence $\{x_n\}$ converge to some point x_0. But then

$$f(x_0)=\varliminf_{x\to x_0} f(x)\leqslant \lim_{n\to\infty} f(x_n)=m,$$

and since the inequality $f(x_0)<m$ is impossible, we must have $f(x_0)=m$.

COROLLARY. *If a function which is lower semi-continuous on a segment does not assume the value* $-\infty$, *then it is bounded below.*

We note that a function which is lower semi-continuous might not have a maximum value. Such a function is, for example,

$$f(x)=\begin{cases}x & \text{for } 0\leqslant x<1,\\ 0 & \text{for } x=1.\end{cases}$$

Theorems 5 and 6 imply that the sum of two functions which are lower semi-continuous on a segment,[18] and also the limit of a monotonically increasing sequence of such functions, are lower semi-continuous functions. In view of the fact that a function which is continuous on a segment is both upper and lower semi-continuous on this segment, we obtain the following result.

THEOREM 9. *The limit of an increasing (decreasing) sequence of continuous functions is a function which is lower (upper) semi-continuous.*

It is evident this theorem is not reversible. In fact, if $f(x)$ equals $-\infty$ at any point, then it cannot be the limit of an *increasing* sequence of continuous functions because the latter are finite everywhere. However, if we restrict ourselves to functions which do not assume the value $-\infty$, then the converse of Theorem 9 holds.

THEOREM 10. *Let a lower semi-continuous function $f(x)$ be defined on $[a,b]$. If $f(x)$ does not take on the value $-\infty$, then there exists an increasing sequence of continuous functions*

$$f_1(x)\leqslant f_2(x)\leqslant f_3(x)\leqslant\dots$$

such that

$$f(x)=\lim_{n\to\infty} f_n(x).$$

Proof. We fix a point $x\in[a,b]$ and consider the function with argument z,

$$f(z)+n|z-x|, \tag{3}$$

[17]Which can equal $-\infty$ (or $+\infty$ if $f(x)\equiv+\infty$).

[18]Under the assumption that this sum exists.

where n is a natural number. The function $|z-x|$ is continuous and, therefore, it is lower semi-continuous. This means that the sum (3) is lower semi-continuous. Since it does not assume the value $-\infty$, it has a minimum value which may be assumed finite.[19]

Denoting this minimum value by $f_n(x)$,

$$f_n(x) = \min\{f(z) + n|z-x|\}, \tag{4}$$

we obtain the required continuous functions.

In fact, let $z_n = z_n(x)$ be one of the points of $[a,b]$ where (3) attains the value $f_n(x)$. Then

$$f_n(x) = f[z_n(x)] + n|z_n(x) - x|. \tag{5}$$

Since $f_n(x)$ is the smallest of the values of function (3), we have that

$$f_n(x) \leqslant f[z_n(y)] + n|z_n(y) - x|$$

for an arbitrary point $y \in [a,b]$. This implies, a fortiori, that

$$f_n(x) \leqslant f[z_n(y)] + n|z_n(y) - y| + n|y - x|.$$

By (5), the last inequality means that $f_n(x) \leqslant f_n(y) + n|y-x|$, from which in view of the interchangeability of the points x and y it follows that

$$|f_n(x) - f_n(y)| \leqslant n|x-y|$$

and, therefore, $f_n(x)$ is a continuous function.

Further, if $m > n$, then

$$f_n(x) \leqslant f[z_m(x)] + n|z_m(x) - x| \leqslant f[z_m(x)] + m|z_m(x) - x| = f_m(x),$$

so that $f_1(x) \leqslant f_2(x) \leqslant f_3(x) \leqslant \ldots$. On the other hand, setting $z = x$ in the inequality

$$f_n(x) \leqslant f(z) + n|z-x|,$$

which holds for all $z \in [a,b]$, we obtain

$$f_n(x) \leqslant f(x). \tag{6}$$

Therefore, setting $g(x) = \lim_{n\to\infty} f_n(x)$, we have

$$g(x) \leqslant f(x). \tag{7}$$

Till now we did not use the continuity of $f(x)$; everything discussed is true for an arbitrary function $f(x)$ which is not identically equal to $+\infty$ and is bounded below.[20]

Now let us momentarily assume that $f(x)$ does not take on the value $+\infty$ anywhere (which, of course, does not at all mean that $f(x)$ is bounded above). Then it follows from (5) and (6) (taking into account that $f(z)$ is bounded below) that $z_n(x) \to x$ as $n \to \infty$. We take an $\varepsilon > 0$; to this ε there corresponds a $\delta > 0$ such that $f(z) > f(x) - \varepsilon$ for $|z-x| < \delta$. For sufficiently large n, $|z_n(x) - x| < \delta$ and, therefore, $f[z_n(x)] > f(x) - \varepsilon$; this means, a fortiori, that

$$f_n(x) > f(x) - \varepsilon.$$

---

[19] If the minimal value were equal to $+\infty$, then this would mean that $f(x) \equiv +\infty$; but then the theorem is trivial because we could have simply set $f_n(x) \equiv n$.

[20] The presence of the minimal value for the function (3) is not essential; if it did not have a minimal value, we could have set

$$f_n(x) = \inf\{f(z) + n|z-x|\}$$

in place of (4). This would only complicate the discussion a little.

This implies that $g(x) > f(x) - \varepsilon$ and since ε is arbitrary, $g(x) \geqslant f(x)$, from which, in combination with (7), the theorem follows.

Now let us drop the assumption that $f(x)$ is continuous and set

$$g_n(x) = \begin{cases} f(x), & \text{if } f(x) \leqslant n, \\ n, & \text{if } f(x) > n. \end{cases}$$

Obviously, $g_1(x) \leqslant g_2(x) \leqslant g_3(x) \leqslant \ldots$ and $\lim g_n(x) = f(x)$. We shall show that the functions $g_n(x)$ are lower semi-continuous. In fact, if $x_0 \in [a, b]$ and $\varepsilon > 0$, then there exists a $\delta > 0$ such that, for $|x - x_0| < \delta$,

$$f(x) > f(x_0) - \varepsilon,$$

and, a fortiori [since $g_n(x_0) \leqslant f(x_0)$], that

$$f(x) > g_n(x_0) - \varepsilon.$$

On the other hand, we have

$$n > g_n(x_0) - \varepsilon.$$

Therefore, taking $x \in (x_0 - \delta, x_0 + \delta)$, we shall have $g_n(x) > g_n(x_0) - \varepsilon$ independently of whether $g_n(x)$ coincides with $f(x)$ or with n. So, $g_n(x)$ is in fact lower semi-continuous. But $g_n(x)$ is finite. This means, by what has already been proved, that for every n there exists a sequence of continuous functions $\varphi_k^{(n)}(x)$ such that

$$\varphi_1^{(n)}(x) \leqslant \varphi_2^{(n)}(x) \leqslant \varphi_3^{(n)}(x) \leqslant \ldots, \lim_{k \to \infty} \varphi_k^{(n)}(x) = g_n(x).$$

We set

$$f_n(x) = \max \{\varphi_n^{(1)}(x), \varphi_n^{(2)}(x), \ldots, \varphi_n^{(n)}(x)\}.$$

It is not difficult to see that[21]

$$f_1(x) \leqslant f_2(x) \leqslant f_3(x) \leqslant \ldots.$$

Furthermore, all the $f_n(x)$ are continuous as was established in Corollary 1 to Lemma 1, §2. Finally, it is easily verified that

$$\lim_{n \to \infty} f_n(x) = f(x),$$

which concludes the proof.

In conclusion, we shall prove an interesting theorem on the "separation of semi-continuous functions by continuous functions."

THEOREM 11. *Let two functions $u(x)$ and $v(x)$ be defined on $[a, b]$, where $u(x)$ is upper semi-continuous, $v(x)$ is lower semi-continuous, and $u(x) \leqslant v(x)$. If $u(x)$ does not take on the value $+\infty$ and $v(x)$ the value $-\infty$. then there exists a continuous function $f(x)$ such that*

$$u(x) \leqslant f(x) \leqslant v(x).$$

Proof. By the preceding theorem, there exist two sequences $\{\varphi_n(x)\}$ and $\{\psi_n(x)\}$ of continuous functions such that

$$\varphi_1(x) \geqslant \varphi_2(x) \geqslant \varphi_3(x) \geqslant \ldots, \qquad \lim_{n \to \infty} \varphi_n(x) = u(x),$$

$$\psi_1(x) \leqslant \psi_2(x) \leqslant \psi_3(x) \leqslant \ldots, \qquad \lim_{n \to \infty} \psi_n(x) = v(x).$$

We introduce the function $f_+(x)$ as usual by setting

---

[21] In fact, let $f_n(x) = \varphi_n^{(k_n)}(x)$. Then

$$f_{n+1}(x) = \max \{\varphi_{n+1}^{(1)}(x), \ldots, \varphi_{n+1}^{(n+1)}(x)\} \geqslant \varphi_{n+1}^{(k_n)}(x) \geqslant \varphi_n^{(k_n)}(x) = f_n(x).$$

$$f_+(x)=\begin{cases}f(x), & \text{if } f(x)\geqslant 0,\\ 0, & \text{if } f(x)<0.\end{cases}$$

Obviously, if $f(x)\leqslant F(x)$, then $f_+(x)\leqslant F_+(x)$. Having noted this, we form the functions

$$(\varphi_1-\psi_1)_+\geqslant(\varphi_1-\psi_2)_+\geqslant(\varphi_2-\psi_2)_+\geqslant(\varphi_2-\psi_3)_+\geqslant\dots$$
$$\dots\geqslant(\varphi_n-\psi_n)_+\geqslant(\varphi_n-\psi_{n+1})_+\geqslant(\varphi_{n+1}-\psi_{n+1})_+\geqslant\dots.$$

Since the difference $\varphi_n(x)-\psi_n(x)$ tends to $u(x)-v(x)\leqslant 0$, we have that $(\varphi_n-\psi_n)_+\to 0$.[22] Therefore, the alternating series

$$f(x)=\psi_1(x)+[\varphi_1(x)-\psi_1(x)]_+-[\varphi_1(x)-\psi_2(x)]_++[\varphi_2(x)-\psi_2(x)]_++\dots \qquad (8)$$

converges everywhere on $[a,b]$.

If $u(x)=v(x)$ at any point x, then $\varphi_i(x)\geqslant\psi_k(x)$ at this point for all i and k. Hence

$$[\varphi_n(x)-\psi_n(x)]_+=\varphi_n(x)-\psi_n(x),\ [\varphi_n(x)-\psi_{n+1}(x)]_+=\varphi_n(x)-\psi_{n+1}(x),$$

and therefore

$$f(x)=\psi_1(x)+\{\varphi_1(x)-\psi_1(x)\}-\{\varphi_1(x)-\psi_2(x)\}+\{\varphi_2(x)-\psi_2(x)\}-\dots.$$

The partial sums of this series are $\psi_1(x),\varphi_1(x),\psi_2(x),\varphi_2(x),\dots$. Hence

$$F(x)=\lim\varphi_n(x)=\lim\psi_n(x)=u(x)=v(x).$$

But if $u(x)<v(x)$ at the point x, then for sufficiently large n, we have $\varphi_n(x)<\psi_n(x)$ [and then, a fortiori, $\varphi_n(x)<\psi_{n+1}(x)$]. Therefore, the terms of series (8), beginning with some one of them, equal zero. If the *first* of the brackets $[\dots]_+$ which is equal to zero is $[\varphi_n(x)-\psi_n(x)]_+$, then

$$f(x)=\psi_1(x)+\{\varphi_1(x)-\psi_1(x)\}-\dots-\{\varphi_{n-1}(x)-\psi_n(x)\}=\psi_n(x).$$

This means that $f(x)\leqslant v(x)$. Furthermore, $\psi_n(x)\geqslant\varphi_n(x)$,[23] which means that $f(x)\geqslant u(x)$. Thus, $u(x)\leqslant f(x)\leqslant v(x)$. This inequality is established analogously if the first of the brackets in series (8) which takes on the value zero has the form $[\varphi_n(x)-\psi_{n+1}(x)]_+$. So, we always have $u(x)\leqslant f(x)\leqslant v(x)$. It remains to show that $f(x)$ is continuous. But all terms of series (8), and hence all its partial sums, are continuous. On the other hand, series (8) is an alternating series, and the absolute value of its general term decreases. In such a series, the partial sums behave so:

$$S_2\leqslant S_4\leqslant S_6\leqslant\dots,\qquad S_1\geqslant S_3\geqslant S_5\geqslant\dots.$$

Therefore, the sum $f(x)$ of series (8) will be both the limit of an increasing sequence of continuous functions and the limit of a decreasing sequence of continuous functions. In other words, $f(x)$ is both lower and upper semi-continuous; it is then continuous. This completes the proof of the theorem.

---

[22] Here we use the fact that the relation $f_n(x)\to F(x)$ implies the relation $[f_n(x)]_+\to F_+(x)$.

[23] Because $[\varphi_n(x)-\psi_n(x)]_+=0$.

CHAPTER XVI

CERTAIN GENERALIZATIONS OF THE LEBESGUE INTEGRAL

§1. INTRODUCTION

At the end of Chapt. IX, we introduced the example

$$f(x) = x^2 \cos \frac{\pi}{x^2} \qquad [f(0) = 0];$$

the function $f(x)$ has a finite derivative $f'(x)$ everywhere on [0, 1], but the latter turns out to be non-integrable in the Lebesgue sense. Thus, the operation of Lebesgue integration does not completely solve the problem of reconstructing the primitive function provided its finite derivative is known. In 1912, the French mathematician A. Denjoy introduced a more general process of integration[1] than that of Lebesgue and showed that this process solves completely the aforementioned problem.

On the other hand, in 1914 the German mathematician O. Perron also introduced a certain definition of integral, based on other principles than the Denjoy definition, which also solved completely the problem of reconstructing the primitive function from its finite derivative.

Later works of H. Hake (1921), P. S. Aleksandrov (1924), and H. Looman (1925) established, however, the identity of the Denjoy and Perron integrals. Thus, it turned out that Perron only proposed a new form of the definition of the Denjoy integral; for this reason, this integral is, at the present time, usually referred to as the Denjoy-Perron integral.

In 1916, independently of each other, A. Denjoy and the Russian mathematician A. Ya. Khinchin gave a still more general definition of integral, which enables one to reconstruct the primitive function not only from its ordinary derivative but also from its so-called approximate (or asymptotic) derivative.[2] This more general integral is usually called the Denjoy-Khinchin integral, or the Denjoy integral "in the wide sense"; in contrast to this, the Denjoy-Perron integral is called the Denjoy integral "in the restricted sense."

We shall study the theory of Denjoy-Perron integrals in detail. As concerns Denjoy-

---

[1]Denjoy himself called his integration process "totalization" and the integral constructed by him a "total," but we shall not adhere to this terminology.

[2]The number A is called the approximate derivative of the function $f(x)$ at the point x_0 if there exists a set E, having x_0 as a density point, such that for $x \in E$ and $x \to x_0$ we have

$$\lim \frac{f(x) - f(x_0)}{x - x_0} = A.$$

[3]S. Saks, *Theory of the Integral*, Monografie Matematyczne, 1937.

Khinchin integrals, we shall restrict ourselves to their definition and then refer the reader for details to the monograph by S. Saks.[3]

§2. DEFINITION OF THE PERRON INTEGRAL

We shall need the following definition in the sequel.

DEFINITION 1. Suppose $F(x)$ is a finite function, defined on the segment $[a,b]$, and that x_0 is a point in this closed interval. The numbers

$$\underline{D}F(x_0)=\underline{\lim}_{x\to x_0}\frac{F(x)-F(x_0)}{x-x_0},\qquad \bar{D}F(x_0)=\overline{\lim}_{x\to x_0}\frac{F(x)-F(x_0)}{x-x_0}$$

are called the *lower* and *upper derivatives* respectively of the function $F(x)$ at the point x_0.

It is easily seen that these numbers (which may be equal to $+\infty$ or $-\infty$) are the least and greatest of the derived numbers of $F(x)$ at the point x_0.

LEMMA 1. *Suppose $U(x)$ and $V(x)$ are two finite functions defined on $[a,b]$. If for some $x_0\in[a,b]$ we have*

$$\underline{D}U(x_0)>-\infty,\qquad \bar{D}V(x_0)<+\infty \tag{1}$$

and $R(x)=U(x)-V(x)$, then

$$\underline{D}R(x_0)\geqslant \underline{D}U(x_0)-\bar{D}V(x_0).$$

Proof. We consider a sequence $\{h_k\}$ such that $h_k\neq 0$, $h_k\to 0$ and

$$\lim\frac{R(x_0+h_k)-R(x_0)}{h_k}=\underline{D}R(x_0).$$

Taking, if necessary, a subsequence of $\{h_k\}$ rather than the sequence $\{h_k\}$ itself, we can attain the situation that the definite limits

$$\lambda=\lim\frac{U(x_0+h_k)-U(x_0)}{h_k},\qquad \mu=\lim\frac{V(x_0+h_k)-V(x_0)}{h_k}$$

exist.

According to (1), we shall have $\lambda>-\infty$ and $\mu<+\infty$, so that the difference $\lambda-\mu$ has meaning. In virtue of this, we have

$$\underline{D}R(x_0)=\lambda-\mu.$$

It remains to note that $\lambda\geqslant \underline{D}U(x_0)$, $\mu\leqslant\bar{D}V(x_0)$.

COROLLARY. *If $U_1(x)$ and $U_2(x)$ are finite functions and*

$$\underline{D}U_1(x)>-\infty,\ \underline{D}U_2(x)>-\infty,$$

then

$$\underline{D}[U_1(x)+U_2(x)]\geqslant \underline{D}U_1(x)+\underline{D}U_2(x).$$

In fact, setting $U_2(x)=-V(x)$ and noting that

$$\underline{D}U_2(x)=-\bar{D}V(x),$$

we can apply the lemma.

DEFINITION 2. Suppose $f(x)$ is a function (which is not necessarily finite), defined on $[a,b]$. The function $F(x)$, *continuous* on $[a,b]$, is said to be:[4]

a *superfunction* for $f(x)$ if	a *subfunction* for $f(x)$ if
1) $F(a)=0$,	1) $F(a)=0$,
2) $\underline{D}F(x)>-\infty$ for all $x\in[a,b]$,	2) $\bar{D}F(x)<+\infty$ for all $x\in[a,b]$,
3) $\underline{D}F(x)\geqslant f(x)$ for all $x\in[a,b]$.	3) $\bar{D}F(x)\leqslant f(x)$ for all $x\in[a,b]$.

[4]This will be our translation of the German terms *Oberfunktion* and *Unterfunktion*.

The concepts of superfunction and subfunction are generalizations of the concept of a primitive function. Namely, we have the following entirely evident lemma.

LEMMA 2. *If the finite function $f(x)$ is the derivative of the function $F(x)$ (where $F(a)=0$), then $F(x)$ is both a superfunction and a subfunction for $f(x)$.*

The next lemma is of great significance for the remainder of this chapter.

LEMMA 3. *If $U(x)$ is a superfunction and $V(x)$ is a subfunction for the same function $f(x)$, then the difference $R(x)=U(x)-V(x)$ is non-decreasing.*

Proof. We have, in virtue of Lemma 1, that

$$\underline{D}\,R(x)\geqslant \underline{D}\,U(x)-\overline{D}\,V(x)\geqslant 0$$

everywhere on $[a,b]$, and everything reduces now to Lemma 1, §7, Chapt. IX.

COROLLARY 1. *Under the conditions of Lemma 3, we have $U(b)\geqslant V(b)$.*

We obtain the next corollary by combining the last proposition with Lemma 2.

COROLLARY 2. *Under the conditions of Lemma 2, the number $F(b)$ is simultaneously the least of the numbers $U(b)$ and the greatest of the numbers $V(b)$, where $U(x)$ and $V(x)$ are arbitrary superfunctions and subfunctions for the function $f(x)$,*

$$F(b)=\min\{U(b)\}=\max\{V(b)\}.$$

We can now state the following fundamental definition.

DEFINITION 3. A function $f(x)$, defined on the closed interval $[a,b]$, is said to be *integrable in the Perron sense* [or to be (P)-integrable] on this segment if

1) it has at least one superfunction $U(x)$ and at least one subfunction $V(x)$;

2) the greatest lower bound of the set $\{U(b)\}$ of the values of all superfunctions at the point $x=b$ coincides with the least upper bound of the set $\{V(b)\}$ of the values of all subfunctions at the same point,

$$\inf\{U(b)\}=\sup\{V(b)\}. \tag{2}$$

If $f(x)$ is (P)-integrable on the closed interval $[a,b]$, then the common value of the numbers (2) is called the *Perron integral* of the function on the closed interval $[a,b]$ and is denoted by

$$(P)\int_a^b f(x)dx.$$

Corollary 2, stated above, to Lemmas 2 and 3 can now be formulated in the form of a theorem:[5]

THEOREM. *If the function $F(x)$ has a finite*[6] *derivative $f(x)$ everywhere on $[a,b]$, then the latter is (P)-integrable and*

$$F(b)-F(a)=(P)\int_a^b f(x)dx.$$

Thus, the Perron integral actually does solve completely the problem of reconstructing the primitive function from its finite derivative. One must not, however, fail to note

---

[5]The existence of P-integrable functions which are not L-integrable follows immediately from this theorem. Such a function is, for example, the derivative of the function $f(x)=x^2\cos\frac{\pi}{x^2}$ $[0<x\leqslant 1]$, $f(0)=0$.

[6]We cannot omit the condition that $f(x)$ is finite. See V. Ya. Kozlov, *Goldovsky's example*, Matem. Sbornik, 1951, 28, No. 1, pp. 197-204.

that the definition introduced above of the Perron integral does not give any process of constructing this integral. This defect will be removed in the theory of the Denjoy integral, which, as we shall see further on, is defined with the aid of a perfectly definite construction.

In concluding this section, we note a simple and important condition for the (P)-integrability of functions, which follows directly from the very definition of the Perron integral.

LEMMA 4. *A necessary and sufficient condition that the function $f(x)$, defined on the closed interval $[a,b]$, be (P)-integrable on this interval is that to every $\varepsilon>0$ there correspond a superfunction $U(x)$ and a subfunction $V(x)$ such that $U(b)-V(b)<\varepsilon$.*

§3. FUNDAMENTAL PROPERTIES OF THE PERRON INTEGRAL

THEOREM 1. *A (P)-integrable function $f(x)$ is finite almost everywhere.*

Proof. Suppose $U(x)$ and $V(x)$ are a superfunction and a subfunction, respectively, for the function $f(x)$, which are (P)-integrable on the closed interval $[a,b]$. We set $R(x)=U(x)-V(x)$. In virtue of Lemma 1, §2, we have that

$$\underline{D}\,R(x)\geqslant \underline{D}\,U(x)-\overline{D}\,V(x)$$

everywhere on $[a,b]$.

But, if $f(x_0)=+\infty$ at some point x_0, then, a fortiori, we have $D\,U(x_0)=+\infty$; since $\overline{D}\,V(x_0)<+\infty$, we have $\underline{D}\,R(x_0)=+\infty$.

The preceding equality also holds at every point x_0 where $f(x_0)=-\infty$. Thus, the set of points at which $f(x)$ assumes the values $\pm\infty$ is a subset of the set of points at which $R'(x)=+\infty$, and the theorem follows from the fact that the derivative of the continuous increasing function $R(x)$ can equal $+\infty$ only on a set of measure 0.

THEOREM 2. *If $f(x)$ is (P)-integrable on the closed interval $[a,b]$ and $a<c<b$, then $f(x)$ is also (P)-integrable on each of the segments $[a,c]$ and $[c,b]$, and*[7]

$$(P)\int_a^b f(x)\,dx=(P)\int_a^c f(x)\,dx+(P)\int_c^b f(x)\,dx. \tag{1}$$

Proof. Suppose $\varepsilon>0$ and that $U(x)$, $V(x)$ are a superfunction and a subfunction respectively for the function $f(x)$ such that

$$U(b)-V(b)<\varepsilon.$$

It is clear that they will also be a superfunction and a subfunction for $f(x)$ on the segment $[a,c]$. On the other hand, the difference $U(x)-V(x)$ increases and hence

$$U(c)-V(c)<\varepsilon.$$

Inasmuch as ε is arbitrary, it follows from this that $f(x)$ is (P)-integrable on the segment $[a,c]$. We note that

$$V(c)\leqslant(P)\int_a^c f(x)\,dx\leqslant U(c). \tag{2}$$

For the segment $[c,b]$,

$$U^*(x)=U(x)-U(c),\ V^*(x)=V(x)-V(c)$$

will serve as a superfunction and subfunction respectively for $f(x)$.

---

[7]Equality (1) means that the Perron integral is an additive function of the interval of integration.

Since

$$U^*(b) - V^*(b) \leqslant U(b) - V(b) < \varepsilon,$$

the function $f(x)$ is also (P)-integrable on $[c, b]$, where

$$V(b) - V(c) \leqslant (P)\int_a^b f(x)dx \leqslant U(b) - U(c). \tag{3}$$

Combining (2) and (3), we find that

$$V(b) \leqslant (P)\int_a^c f(x)dx + (P)\int_c^b f(x)dx \leqslant U(b),$$

from which (1) follows.

THEOREM 3. *If $a < c < b$ and $f(x)$ is (P)-integrable on each of the segments $[a, c]$ and $[c, b]$, then it is also (P)-integrable on all of $[a, b]$.*

Proof. Suppose $\varepsilon > 0$ and that $U_1(x)$, $V_1(x)$ are a superfunction and subfunction respectively for $f(x)$ on $[a, c]$, and that $U_2(x)$, $V_2(x)$ are a superfunction and subfunction respectively for $f(x)$ on $[c, b]$, where

$$U_1(c) - V_1(c) < \varepsilon, \quad U_2(b) - V_2(b) < \varepsilon.$$

We introduce the functions

$$U(x) = \begin{cases} U_1(x) & \text{for } a \leqslant x \leqslant c, \\ U_1(c) + U_2(x) & \text{for } c \leqslant x \leqslant b, \end{cases}$$

$$V(x) = \begin{cases} V_1(x) & \text{for } a \leqslant x \leqslant c, \\ V_1(c) + V_2(x) & \text{for } c \leqslant x \leqslant b. \end{cases}$$

It is easily seen[8] that $U(x)$, $V(x)$ are a superfunction and subfunction respectively for the function $f(x)$ on all of $[a, b]$. Inasmuch as $U(b) - V(b) < 2\varepsilon$, and ε is arbitrary, the theorem is proved.

THEOREM 4. *If $f(x)$ is (P)-integrable on $[a, b]$ and k is a finite constant, then the function $kf(x)$ is also (P)-integrable on $[a, b]$ and*

$$(P)\int_a^b kf(x)dx = k(P)\int_a^b f(x)dx. \tag{4}$$

Proof. For $k = 0$, the theorem is trivial.[9] Suppose $k > 0$. If $U(x)$ and $V(x)$ are a superfunction and subfunction for $f(x)$, then[10] $kU(x)$, $kV(x)$ are respectively a superfunction and subfunction for $kf(x)$. In view of the fact that the difference $kU(b) - kV(b)$, as well as $U(b) - V(b)$, can be made arbitrarily small, the function $kf(x)$ will be (P)-integrable. Equality (4) follows from the inequalities

$$kV(b) \leqslant (P)\int_a^b kf(x)dx \leqslant kU(b).$$

---

[8] In fact, $\underline{D}U(c) = \min\{\underline{D}U_1(c), \underline{D}U_2(c)\}$, from which both of the inequalities $\underline{D}U(c) > -\infty$ and $\underline{D}U(c) \geqslant f(c)$ follow.

[9] The function $\varphi(x) \equiv 0$ is the derivative of the function $F(x) \equiv c$. This means, according to the theorem in §2, that $\varphi(x)$ is P-integrable and that $(P)\int_a^b 0dx = 0$.

[10] We have $\underline{D}[kF(x)] = k\underline{D}F(x)$ and $\bar{D}[kF(x)] = k\bar{D}F(x)$ for $k > 0$.

Finally, let $k<0$. Then $kV(x)$, $kU(x)$ will be a superfunction and subfunction respectively for $kf(x)$ and the proof is completed as above.

THEOREM 5. *Let each of the functions $f_1(x)$ and $f_2(x)$ be (P)-integrable on the closed interval $[a,b]$. If their sum is defined on all of $[a,b]$, then it is also (P)-integrable on this segment and*

$$(P)\int_a^b [f_1(x) + f_2(x)]dx = (P)\int_a^b f_1(x)dx + (P)\int_a^b f_2(x)dx.$$

Proof. Suppose $U_1(x)$ and $U_2(x)$ are superfunctions for $f_1(x)$, $f_2(x)$ respectively. If $U(x)=U_1(x)+U_2(x)$, then this function will be a superfunction for the sum $f_1(x)+f_2(x)$ [see Corollary to Lemma 1, §2]. Analogously, the sum of the subfunctions of the functions $f_1(x)$ and $f_2(x)$ will be a subfunction for $f_1(x)+f_2(x)$. The remainder of the proof is clear.

THEOREM 6. *If the function $f(x)$ is (P)-integrable on the segment $[a,b]$, and the function $g(x)$, defined on the same segment, is equivalent to it, then $g(x)$ is also (P)-integrable on $[a,b]$ and*

$$(P)\int_a^b g(x)dx = (P)\int_a^b f(x)dx. \tag{5}$$

Proof. We denote the set of points at which $g(x)\neq f(x)$ by E.

Since $mE=0$, there exists a continuous increasing function $\sigma(x)$ such that $\sigma'(x)=+\infty$ for all points of E (see Theorem 6, §2, Chapt. VIII).

We can assume[11] that $\sigma(a)=0$, $\sigma(b)=1$.

We take $\varepsilon>0$ and consider a superfunction $U(x)$ and a subfunction $V(x)$ for the function $f(x)$ such that

$$U(b)-V(b)<\varepsilon.$$

Suppose

$$U^*(x)=U(x)+\varepsilon\sigma(x), \quad V^*(x)=V(x)-\varepsilon\sigma(x).$$

It is easy to show[12] that $U^*(x)$ and $V^*(x)$ are already a superfunction and subfunction respectively for $g(x)$. But since

$$U^*(b)-V^*(b)<3\varepsilon,$$

we have that $g(x)$ is (P)-integrable. Finally, the inequalities

$$V(b)-\varepsilon\leqslant(P)\int_a^b g(x)dx\leqslant U(b)+\varepsilon$$

imply (5).

---

[11]Otherwise we must consider

$$\sigma_1(x)=\frac{\sigma(x)-\sigma(a)}{\sigma(b)-\sigma(a)}$$

instead of $\sigma(x)$.

[12]In fact, the fact that $\sigma(x)$ is increasing implies that $\underline{D}U^*(x)\geqslant\underline{D}U(x)$. Therefore we have $\underline{D}U^*(x)>-\infty$ everywhere on $[a,b]$. Now it follows from this that $\underline{D}U^*(x)\geqslant g(x)$ for $x\overline{\in}E$. However, for $x\in E$, the last inequality is obvious since $\underline{D}U^*(x)\geqslant\underline{D}U(x)+\underline{D}\sigma(x)=+\infty$ for such x. Thus, $U^*(x)$ is a superfunction for $g(x)$. The reasoning is analogous for $V^*(x)$.

Remark. Theorem 6 enables one to omit, in the formulation of Theorem 5, the agreement that the sum $f_1(x)+f_2(x)$ must be defined at all points of $[a,b]$. In fact, since the terms are almost everywhere finite (by Theorem 1), the sum $f_1(x)+f_2(x)$ is defined, in every case, almost everywhere on $[a,b]$, and it can be defined arbitrarily on the remaining set of measure 0.

§4. THE INDEFINITE PERRON INTEGRAL

If $f(x)$ is defined and (P)-integrable on the closed interval $[a,b]$, then the function

$$F(x)=C+(P)\int_a^x f(t)\,dt \qquad (a< x\leqslant b)$$

is called the *indefinite Perron integral* of $f(x)$. If we set

$$(P)\int_a^a f(t)\,dt=0,$$

we can consider $F(x)$ for all x in $[a,b]$.

THEOREM 1. *The indefinite Perron integral of $f(x)$ is continuous.*

Proof. Suppose $U(x)$ and $V(x)$ are a superfunction and subfunction respectively for $f(x)$ satisfying the inequality

$$U(b)-V(b)<\varepsilon.$$

Since the difference $U(x)-V(x)$ increases, we have

$$U(x_0)-V(x_0)<\varepsilon$$

for every x_0 in $[a,b]$.

On the other hand, the functions $U(x)$ and $V(x)$ will also be a superfunction and subfunction respectively for $f(x)$ on every segment $[a,x_0]$. Hence

$$V(x_0)\leqslant(P)\int_a^{x_0} f(t)\,dt\leqslant U(x_0).$$

Thus, for all x in $[a,b]$, we have

$$0\leqslant U(x)-(P)\int_a^x f(t)\,dt<\varepsilon.$$

The possibility of effecting an arbitrarily close uniform approximation to the integral $(P)\int_a^x f(t)dt$ with the aid of the *continuous* functions $U(x)$ now completes the proof of the theorem.

THEOREM 2. *If $f(x)$ is (P)-integrable on $[a,b]$,*

$$F(x)=(P)\int_a^x f(t)\,dt, \tag{1}$$

and $U(x)$, $V(x)$ are a superfunction and subfunction for $f(x)$, then each of the differences

$$U(x)-F(x), \qquad F(x)-V(x) \tag{2}$$

is an increasing function.

Proof. Suppose $a \leqslant x_1 < x_2 \leqslant b$. The function $U(x) - U(x_1)$ will be a superfunction for $f(x)$ on the segment $[x_1, x_2]$. It follows that

$$(P)\int_{x_1}^{x_2} f(t)\,dt \leqslant U(x_2) - U(x_1),$$

or, equivalently,

$$F(x_2) - F(x_1) \leqslant U(x_2) - U(x_1),$$

and hence

$$U(x_1) - F(x_1) \leqslant U(x_2) - F(x_2).$$

The second difference in (2) is treated analogously.

THEOREM 3. *A function which is (P)-integrable almost everywhere is the derivative of its indefinite Perron integral.*

Proof. Retaining the notation in (1), we can find a superfunction $U(x)$ such that

$$U(b) - F(b) < \varepsilon^2,$$

where $\varepsilon > 0$ is a prescribed number. Suppose

$$R(x) = U(x) - F(x).$$

This function is increasing and continuous. As is known, it has a finite derivative $R'(x)$ almost everywhere and

$$(L)\int_a^b R'(x)\,dx \leqslant R(b) - R(a) < \varepsilon^2.$$

We denote the set of those points for which

$$\underline{D}F(x) < f(x) - \varepsilon$$

by $A(\varepsilon)$.

For each of these points, we shall have a fortiori that $\underline{D}F(x) < \underline{D}U(x) - \varepsilon$, from which it follows[13] that $\underline{D}U(x) - \underline{D}F(x) > \varepsilon$.

Further, we denote the set of those points in $[a, b]$ where $R'(x)$ exists and is finite by M (thus, $mM = b - a$). If $x \in M$, then [14]

$$\underline{D}U(x) - \underline{D}F(x) = R'(x).$$

Hence, the intersection $MA(\varepsilon)$ is a subset of the set $B(\varepsilon)$ of those points of $[a, b]$ at which $R'(x) > \varepsilon$. But

$$\varepsilon m B(\varepsilon) \leqslant (L)\int_{B(\varepsilon)} R'(x)\,dx \leqslant (L)\int_a^b R'(x)\,dx < \varepsilon^2,$$

from which it follows that $mB(\varepsilon) < \varepsilon$ and hence $m^*A(\varepsilon) < \varepsilon$.

Replacing ε by $\dfrac{\varepsilon}{2^n}$ in the last inequality, we find that

$$m^*A\left(\frac{\varepsilon}{2^n}\right) < \frac{\varepsilon}{2^n}.$$

---

[13] The inequality $\underline{D}F(x) < f(x) - \varepsilon$ guarantees that $\underline{D}F(x) < +\infty$ and since $\underline{D}U(x) > -\infty$, the difference $\underline{D}U(x) - \underline{D}F(x)$ has meaning.

[14] In fact, $\dfrac{U(x+h) - U(x)}{h} = \dfrac{F(x+h) - F(x)}{h} + \dfrac{R(x+h) - R(x)}{h}$. From this it is clear that $\underline{D}U(x) = \underline{D}F(x) + R'(x)$.

But obviously

$$m^*E(\underline{D}F < f) < \varepsilon.$$

Since ε is arbitrary, it is clear that

$$\underline{D}F(x) \geqslant f(x) \tag{3}$$

almost everywhere on $[a, b]$.

Introducing a subfunction, it can be proved analogously that

$$\overline{D}F(x) \leqslant f(x) \tag{4}$$

almost everywhere on $[a, b]$; and everywhere where both inequalities (3) and (4) are satisfied, the derivative $F'(x)$ exists and equals $f(x)$.

COROLLARY 1. *Every (P)-integrable function is measurable.*

In fact, retaining the notation introduced above and setting

$$F(x) = F(b) \text{ for } x > b,$$

we see that $f(x)$ can be represented in the form of the limit of a sequence of continuous functions,

$$f(x) = \lim_{n\to\infty} n\left[F\left(x + \frac{1}{n}\right) - F(x)\right],$$

for almost all x in $[a, b]$.

COROLLARY 2. *A superfunction $U(x)$ and subfunction $V(x)$ of an arbitrary (P)-integrable function are differentiable almost everywhere.*

In fact, using the same notation as above, we have

$$U(x) = [U(x) - F(x)] + F(x),$$

and each of the functions $U(x) - F(x)$ and $F(x)$ is differentiable almost everywhere. The situation is analogous for $V(x)$.

§5. COMPARISON OF THE PERRON AND LEBESGUE INTEGRALS

As we have already noted in §2 (see footnote [5], §2 above), there exist functions which are integrable in the Perron sense but are not integrable in the Lebesgue sense.

In this section, we shall show that the converse cannot hold. To this end, we need two lemmas.

LEMMA 1. *Suppose the function $u(x)$ is summable on $[a, b]$ and that*

$$U(x) = (L)\int_a^x u(t)\,dt.$$

If $u(x)$ is lower semi-continuous at the point x_0 then

$$\underline{D}U(x_0) \geqslant u(x_0).$$

Proof. We may assume that $u(x_0) > -\infty$ because otherwise there is nothing to prove. Taking any A which satisfies the inequality $A < u(x_0)$, we can guarantee the existence of a $\delta > 0$ such that $u(t) > A$ for all $t \in [a, b]$ for which $|t - x_0| < \delta$.

In this case, for $|x - x_0| < \delta$, we have[15]

$$(L)\int_{x_0}^{x} u(t)\,dt \geqslant A(x - x_0),$$

from which it follows that

[15] We have adapted the discussion to the case $x > x_0$. The reader can carry out the discussion for $x < x_0$.

$$\frac{U(x)-U(x_0)}{x-x_0} \geqslant A$$

and hence all the derived numbers $DU(x_0)$ are not less than A. In particular, we also have that $\underline{D}U(x_0) \geqslant A$. The remainder of the proof is clear.

LEMMA 2. *Suppose $f(x)$ is integrable on $[a,b]$. Then, for every $\varepsilon>0$, there exists a function $u(x)$ with the following properties:*

1) *it is defined on $[a,b]$ and everywhere lower semi-continuous;*
2) *$u(x)>-\infty$ everywhere;*
3) *$u(x)\geqslant f(x)$ everywhere;*
4) *the function $u(x)$ is summable and*

$$(L)\int_a^b u(x)\,dx<\varepsilon+(L)\int_a^b f(x)\,dx.$$

Proof. I. We shall first assume that $f(x)$ is non-negative and bounded,

$$0\leqslant f(x)< M.$$

We set

$$\sigma=\frac{\varepsilon}{1+b-a}, \tag{1}$$

where ε is the number referred to in the condition of the lemma, and we choose a natural number n so large that $n\sigma> M$.

We consider the sets

$$E_k=E(k\sigma\leqslant f<(k+1)\sigma) \qquad (k=0,1,\ldots,n-1).$$

For each of them, there exists a bounded open set G_k such that

$$G_k\supset E_k,\ mG_k<mE_k+\frac{1}{(k+1)2^{k+1}}.$$

Suppose $S_k=[a,b]G_k$ and that $\varphi_k(x)$ is the characteristic function of the set S_k (we define $\varphi_k(x)$ on the initial segment $[a,b]$). All the functions $\varphi_k(x)$ are lower semi-continuous.[16]

We set

$$u(x)=\sigma\sum_{k=0}^{n-1}(k+1)\varphi_k(x)$$

and show that this is the required function. In fact, it is non-negative, lower semi-continuous, and bounded above. Suppose $x\in[a,b]$. Then for some i we have that $x_i\in E_i\subset S_i$ and $\varphi_i(x)=1$. In fact,

$$u(x)\geqslant(i+1)\sigma>f(x).$$

Finally,

$$(L)\int_a^b u(x)\,dx=\sigma\sum_{k=0}^{n-1}(k+1)mS_k<\sigma\sum_{k=0}^{n-1}(k+1)\left[mE_k+\frac{1}{(k+1)2^{k+1}}\right].$$

It follows that

$$(L)\int_a^b u(x)\,dx<\sum_{k=0}^{n-1}\sigma kmE_k+\sigma[(b-a)+1].$$

[16] See the example given in §4, Chapt. XV.

Taking (1) and the inequality

$$\sigma k m E_k \leqslant (L)\int_{E_k} f(x)\,dx$$

into consideration, we find that

$$(L)\int_a^b u(x)\,dx < \varepsilon + (L)\int_a^b f(x)\,dx.$$

Thus, the lemma is proved for the case when $f(x)$ is non-negative and bounded.

II. We shall now omit the assumption that the function $f(x)$ is bounded, but we shall still assume, however, that it is non-negative.

We set

$$S_n(x) = \begin{cases} f(x), & \text{if } f(x) \leqslant n, \\ n, & \text{if } f(x) > n, \end{cases}$$

and suppose

$$f_1(x) = S_1(x), \qquad f_n(x) = S_n(x) - S_{n-1}(x) \qquad (n = 2, 3, \ldots).$$

The functions $f_n(x)$ are measurable, non-negative and bounded, and

$$f(x) = \sum_{n=1}^{\infty} f_n(x).$$

According to what we have already proved, for every natural number n there exists a lower semi-continuous function $u_n(x)$ satisfying the inequalities

$$u_n(x) \geqslant f_n(x), \quad (L)\int_a^b u_n(x)\,dx < \frac{\varepsilon}{2^n} + (L)\int_a^b f_n(x)\,dx.$$

It is easily seen[17] that the function

$$u(x) = \sum_{n=1}^{\infty} u_n(x)$$

satisfies all the requirements of the lemma.

III. Finally, we consider the general case when $f(x)$ is an arbitrary summable function.

We set

$$f_n(x) = \begin{cases} f(x), & \text{if } f(x) \geqslant -n, \\ -n, & \text{if } f(x) < -n. \end{cases}$$

Clearly,

$$|f_n(x)| \leqslant |f(x)|, \qquad \lim_{n\to\infty} f_n(x) = f(x).$$

Hence

$$\lim_{n\to\infty} (L)\int_a^b f_n(x)dx = (L)\int_a^b f(x)dx.$$

We take a fixed n for which

$$(L)\int_a^b f_n(x)\,dx < \frac{\varepsilon}{2} + (L)\int_a^b f(x)\,dx.$$

[17] See the corollary to Theorem 6, §4, Chapt. XV.

The function $n+f_n(x)$ is non-negative and summable. According to what we have already proved, there exists a lower semi-continuous function $u^*(x)$ such that

$$u^*(x) \geqslant n+f_n(x),\ (L)\int_a^b u^*(x)\,dx < \frac{\varepsilon}{2}+(L)\int_a^b [n+f_n(x)]\,dx.$$

Then the function $u(x)=u^*(x)-n$ possesses all the required properties.

THEOREM 1. *If the function $f(x)$ is Lebesgue-integrable on the segment $[a,b]$, then it is also Perron-integrable on this segment and*

$$(P)\int_a^b f(x)\,dx = (L)\int_a^b f(x)\,dx.$$

Proof. We take $\varepsilon>0$ and consider the lower semi-continuous function $u(x)$ which was discussed in the preceding lemma. If

$$U(x)=(L)\int_a^x u(t)\,dt,$$

then $U(x)$ will be a superfunction for $f(x)$. In fact, the continuity of $U(x)$ and the equality $U(a)=0$ are clear. Further, according to Lemma 1, for arbitrary x in $[a,b]$ we have

$$\underline{D}U(x)\geqslant u(x),$$

which implies the two inequalities

$$\underline{D}U(x)>-\infty,\ \underline{D}U(x)\geqslant f(x),$$

which characterize a superfunction.

Inasmuch as the number ε is arbitrary, the inequality

$$U(b)=(L)\int_a^b u(t)\,dt < \varepsilon+(L)\int_a^b f(x)\,dx$$

implies that

$$\inf\{U(b)\} \leqslant (L)\int_a^b f(x)\,dx,$$

where $\{U(b)\}$ is the set of values of all possible superfunctions at $x=b$.

It is established in an analogous manner that

$$\sup\{V(b)\} \geqslant (L)\int_a^b f(x)\,dx,$$

where the notation used is self-evident. But since we always have

$$\sup\{V(b)\} \leqslant \inf\{U(b)\},$$

both these numbers are equal to the integral

$$(L)\int_a^b f(x)\,dx,$$

which completes the proof of the theorem.

The next theorem is an interesting extension of this theorem.

THEOREM 2. *Every non-negative (P)-integrable function is necessarily (L)-integrable.*

Proof. Suppose $f(x)$ is the function under consideration and that $U(x)$ is a super-function for $f(x)$. Then the inequalities

$$\underline{D}U(x) \geqslant f(x) \geqslant 0$$

imply that $U(x)$ is an increasing function and hence its derivative $U'(x)$ (which exists almost everywhere) is summable.[18] But then the theorem follows from the inequalities

$$0 \leqslant f(x) \leqslant U'(x),$$

which hold at all points where $U'(x)$ exists.

COROLLARY. *If $f(x)$ is a measurable function and the Perron integral*

$$(P)\int_a^b |f(x)|\,dx$$

exists, then $f(x)$ is summable.

Thus, the Perron integral is absolutely convergent if, and only if, it reduces to the Lebesgue integral.

§6. ABSTRACT DEFINITION OF THE INTEGRAL AND ITS GENERALIZATIONS

In this section, we shall consider certain general concepts which will be used in the sequel in the discussion of the theory of the Denjoy integral.

We are already acquainted with the R (Riemann), L (Lebesgue), and P (Perron) integrals. All these integrals possess certain general properties and we wish to include these properties in some unifying scheme.

Suppose each segment $[a, b]$ where $a \leqslant b$ is assigned some non-void class $T([a, b])$, consisting of functions defined on $[a, b]$.

We shall agree to call such a system of classes a *legitimate* system if[19]

$$T([a, b]) = T([a, c])\,T([c, b])$$

for arbitrary $c \in [a,b]$.

Suppose $\mathfrak{M} = \{T([a, b])\}$ is any legitimate system of classes and that a functional

$$\mathbf{T}_a^b(f)$$

is defined on each class $T([a, b])$, which assigns to each function f in $T([a, b])$ a definite real number. This functional will be called an integral if for arbitrary $f \in T([a, b])$ and any $c \in [a, b]$ we have[20]

---

[18]See Theorem 5, §2, Chapt. VIII.

[19]The precise meaning of this relation is: the function $f(x)$, defined on $[a, b]$, occurs in $T([a, b])$ if, and only if, both functions obtained from $f(x)$ by considering it only on $[a, c]$ or on $[c, b]$, occur in the corresponding classes $T([a, c])$ and $T([c, b])$.

[20]In particular, if $f \in T([a, a])$, then

$$\mathbf{T}_a^a(f) = \mathbf{T}_a^a(f) + \mathbf{T}_a^a(f),$$

i.e. $\mathbf{T}_a^a(f) = 0$: the integral is "pointwise" equal to zero.

$$\overset{b}{\underset{a}{\mathbf{T}}}(f)=\overset{c}{\underset{a}{\mathbf{T}}}(f)+\overset{b}{\underset{c}{\mathbf{T}}}(f) \tag{1}$$

and (for $x\in[a,b]$)

$$\lim_{x\to c}\overset{x}{\underset{a}{\mathbf{T}}}(f)=\overset{c}{\underset{a}{\mathbf{T}}}(f). \tag{2}$$

In other words, an integral is an additive, continuous interval function.

All functions which occur in the class $T([a,b])$ will be called T-integrable on $[a,b]$. It follows from the legitimacy condition of the system of classes $T([a,b])$ that every function which is T-integrable on the segment $[a,b]$ is also T-integrable on every segment[21] $[p,q]$ which is contained in $[a,b]$.

We shall now consider certain properties of the abstract concept of integral which we have just introduced. Suppose $f(x)$ is a function defined on $[a,b]$ and that $c\in[a,b]$. If for arbitrary $\delta>0$ the function $f(x)$ is not T-integrable on the segment $[c-\delta, c+\delta]\,[a,b]$, then we shall agree to call the point c a T-singularity for the function $f(x)$ and the set of all such points will be denoted by $S_T(f;[a,b])$. Sometimes, instead of $S_T(f;[a,b])$ we shall write $S_T([a,b])$ or $S_T(f)$ or even simply S_T.

It is clear that $S_T([a,b])=0$ for a function which is T-integrable on $[a,b]$. The following lemma shows that the converse is also valid.

LEMMA 1. *If $f(x)$ is defined on $[a,b]$ and does not occur in $T([a,b])$, then*

$$S_T(f;[a,b])\neq 0.$$

Proof. We set $d=\dfrac{a+b}{2}$. At least one of the segments $[a,d]$ and $[d,b]$ has the property that $f(x)$ is not T-integrable on it. We shall denote it by $[a_1,b_1]$ and, setting $d_1=\dfrac{a_1+b_1}{2}$, we denote by $[a_2b_2]$ one of the segments $[a_1,d_1]$ and $[d_1,b_1]$ on which $f(x)$ is not T-integrable. Continuing this process, we construct an infinite sequence of nested segments

$$[a,b]\supset[a_1,b_1]\supset[a_2,b_2]\supset\ldots,$$

on each of which $f(x)$ is not T-integrable.

Suppose c is the point common to all the $[a_n,b_n]$. If $\delta>0$, then $[a_n,b_n]\subset[c-\delta, c+\delta]\,[a,b]$ for sufficiently large n.

From this it follows that $f\in T([c-\delta,c+\delta]\,[a,b])$, and the point c is T-singular.

LEMMA 2. *The set $S_T=S_T(f;[a,b])$ is closed.*

Proof. Suppose $c_n\in S_T$ and that $c_n\to c$. We take $\delta>0$. If n is sufficiently large, then $|c_n-c|<\dfrac{\delta}{2}$ and hence

$$\left[c_n-\frac{\delta}{2},c_n+\frac{\delta}{2}\right][a,b]\subset[c-\delta,c+\delta]\,[a,b].$$

Since $f(x)$ is not T-integrable on the segment in the left member of this inclusion relation, it is a fortiori not T-integrable on $[c-\delta,c+\delta]\,[a,b]$. Inasmuch as δ is arbitrary, it follows from this that $c\in S_T$, which was required to be proved.

In the sequel, we shall be interested in the cases when S_T does not fill the entire

[21] And, in particular on every point $c\in[a,b]$.

segment $[a, b]$. In these cases, the complementary set $[a, b] - S_T$ will consist of a finite or denumerable set of pairwise disjoint intervals. In fact, if $S_T = 0$, then $[a, b] - S_T = [a, b]$. But if $S_T \neq 0$ and $[p, q]$ is the smallest segment containing S_T, then[22]

$$[a, b] - S_T = [a, p) + \{[p, q] - S_T\} + (q, b]$$

and it remains to mention that the set $[p, q] - S_T$ is either void or it consists of pairwise disjoint intervals. The situation where certain of the intervals, complementary to S_T, are not open intervals, strictly speaking, requires corresponding complications in notation, but nonetheless we shall denote all these intervals simply by (a_n, b_n) in the sequel, although it is not excluded that one of the intervals (a_n, b_n) is in fact $[a_n, b_n)$ or $(a_n, b_n]$ and even $[a_n, b_n]$ (if $S_T = 0$).

We shall assume that two integrals T_1 and T_2 are under consideration. If every function which is T_1-integrable is necessarily T_2-integrable, where the values of both integrals coincide for the same function, then we say that the integral T_2 is *more general* than T_1.

We shall show how, starting with some integral T (in the very definition of which the corresponding legitimate system of classes of T-integrable functions $T([a, b])$ must be indicated), one can construct another more general integral T_*.

To this end, we must first of all define the legitimate system of classes of integrable functions corresponding to the new integral. We shall agree to include the function $f(x)$, defined on the segment $[a, b]$, in the class $T_*([a, b])$, if and only if the following three conditions are satisfied:

1) The set $S_T = S_T(f; [a, b])$ is nowhere dense on $[a, b]$ and the function $f(x)$ is Lebesgue-integrable[23] on this set;

2) If $\{(a_n, b_n)\}$ is the set of intervals complementary to S_T, then a finite limit

$$I_n = \lim_{\alpha}^{\beta} \mathbf{T}(f) \qquad (a_n < \alpha < \beta < b_n,\ \alpha \to a_n, \beta \to b_n)$$

exists for every n;

3) If[24]

$$W_n = \sup |\overset{\beta}{\underset{\alpha}{\mathbf{T}}}(f)| \qquad (a_n < \alpha < \beta < b_n),$$

then

$$\sum_n W_n < +\infty.$$

We shall verify that the system of classes $T_*([a, b])$ is legitimate. We shall first assume that $f(x) \in T([a, b])$. Then $S_T(f; [a, b]) = 0$ and the entire sum $\sum(a_n, b_n)$ reduces to one term: the segment $[a, b]$. We have already noted that condition 1) is satisfied by our function (see footnote [23]). Condition 2) is also satisfied for this function because, in virtue of (2), we have

$$\lim_{\alpha}^{\beta} \mathbf{T}(f) = \overset{b}{\underset{a}{\mathbf{T}}}(f),$$

---

[22] For $p = a$, we have $[a, p) = 0$.

[23] This condition is always satisfied when $mS_T = 0$ and a fortiori when $S_T = 0$.

[24] Condition 2) assures the finiteness of the numbers W_n.

if $a<\alpha<\beta<b$, $\alpha\to a$, $\beta\to b$. Finally, condition 3) is also satisfied for $f(x)$ since there is only one (finite !) number W.

Thus,

$$T([a,b])\subset T_*([a,b])$$

and all the classes $T_*([a,b])$ are non-void.

Suppose, further, that $f(x)\in T_*([a,b])$ and[25] that $a<c<b$. We shall consider the set $S_T(f;[a,c])$. It is easily seen that it is a subset of the set $S_T(f;[a,b])$ and hence is it nowhere dense on $[a,c]$, and the function $f(x)$ is Lebesgue-integrable on it. Thus, the function $f(x)$ satisfies condition 1) on the segment $[a,c]$. Suppose

$$[a,c]-S_T(f;[a,c])=\sum_n(a_n,b_n).$$

If $c\in S_T(f;[a,c])$, then every interval which appears here will also be the complement of the entire set $S_T(f;[a,b])$ with respect to the segment $[a,b]$. But if $c\bar{\in}S_T(f;[a,c])$, then this is the situation for all (a_n,b_n) except, perhaps, for one interval of the form[26] $(a_{n_0},c]$. It follows directly from this situation that conditions 2) and 3), which are necessary that $f(x)$ be included in the class $T_*([a,c])$, are satisfied. It is proved analogously that $f\in T_*([c,b])$. Thus,

$$T_*([a,b])\subset T_*([a,c])\,T_*([c,b]).$$

The reverse inclusion is established with the aid of similar reasoning.

Thus, the system of classes $T_*([a,b])$ is legitimate. We shall now define for every $f\in T_*([a,b])$ the value of the functional $\overset{b}{\underset{a}{\mathbf{T}}}_*(f)$ with the aid of the formula

$$\overset{b}{\underset{a}{\mathbf{T}}}_*(f)=\sum_n I_n+(L)\int\limits_{S_T(f)}f(x)\,dx. \tag{3}$$

This definition is meaningful since $|I_n|\leqslant W_n$ and hence the series $\sum I_n$ converges absolutely.

It is useful to note that for every T-integrable function $f(x)$ (we have already noted above that it also occurs in $T_*([a,b])$) we have

$$\overset{b}{\underset{a}{\mathbf{T}}}_*(f)=\overset{b}{\underset{a}{\mathbf{T}}}(f),$$

because the Lebesgue integral vanishes in the right member of (3), and the sum $\sum I_n$ reduces to the single term $\overset{b}{\underset{a}{\mathbf{T}}}(f)$.

We have another, simpler case when a and b are the only T-singular points on $[a,b]$. In this case,

$$\overset{b}{\underset{a}{\mathbf{T}}}_*(f)=\lim\overset{\beta}{\underset{\alpha}{\mathbf{T}}}(f)\qquad(a<\alpha<\beta<b,\quad\alpha\to a,\beta\to b).$$

In virtue of this simple remark, it follows from (3) that

---

[25]The cases $c=a$ and $c=b$ are trivial. In fact if $f(x)$ is defined at the point x_0 then, independently of whether or not it is T-integrable at this point, $f(x)$ necessarily occurs in the class $T_*([x_0,x_0])$.

[26]Moreover, it is possible that $S_T(f;[a,c])=0$. Then the exclusion referred to will be the segment $[a,c]$ itself.

$$\overset{b}{\underset{a}{\mathbf{T}}}_*(f) = \sum_n \overset{b_n}{\underset{a_n}{\mathbf{T}}}_*(f) + (L)\int_{S_T(f)} f(x)\,dx.$$

We shall verify that the functional $\overset{b}{\underset{a}{\mathbf{T}}}_*(f)$ which we defined above is an *integral* in the sense of the definition given at the beginning of this section, i.e. that it is additive and continuous as a function of the interval $[a, b]$.

Suppose $f \in T_*([a, b])$ and that $a < c < b$. It is clear that

$$S_T(f; [a, b]) = S_T(f; [a, c]) + S_T(f; [c, b]),$$

where the sets written in the right either do not intersect at all or they have only one point in common. It follows from this that

$$(L)\int_{S_T([a,\,b])} f(x)\,dx = (L)\int_{S_T([a,\,c])} f(x)\,dx + (L)\int_{S_T([c,\,b])} f(x)\,dx, \tag{4}$$

where both integrals in the right member exist and are finite.

Further, suppose

$$[a, b] - S_T(f; [a, b]) = \sum_n (a_n, b_n).$$

Three assumptions can be made concerning the point c: it occurs in both the sets $S_T([a, c])$ and $S_T([c, b])$; it occurs in neither of these sets; it occurs in one and not the other. In all these cases, the reasoning is very similar and, as an example, we shall consider only the first case. In this case, the entire set of intervals (a_n, b_n) decomposes into two disjoint subsets consisting of intervals situated respectively to the left and to the right of the point c. Using self-evident notation, we have

$$\sum_{[a,\,b]} I_n = \sum_{[a,\,c]} I_n + \sum_{[c,\,b]} I_n,$$

from which it follows, in connection with (4), that[27]

$$\overset{b}{\underset{a}{\mathbf{T}}}_*(f) = \overset{c}{\underset{a}{\mathbf{T}}}_*(f) + \overset{b}{\underset{c}{\mathbf{T}}}_*(f). \tag{5}$$

Thus, the function $\mathbf{T}_*$ is an additive interval function. We shall now prove that for $f \in \mathbf{T}_*([a, b])$, $x \in [a, b]$, $c \in [a, b]$ and $x \to c$ we have

$$\lim \overset{x}{\underset{a}{\mathbf{T}}}_*(f) = \overset{c}{\underset{a}{\mathbf{T}}}_*(f). \tag{6}$$

Suppose, for definiteness, that $x < c$. Then we may assume that $c = b$. We can make three assumptions concerning the point b: it does not occur in the set $S_T(f; [a, b])$, it is an isolated point of this set, or it is a limit point of this set.

In the first two cases, the reasoning is almost evident. In fact, if $b \bar{\in} S_T(f; [a, b])$, then the function $f(x)$ will be T-integrable on the segment $[p, b]$, provided p is sufficiently close to b. Hence, for such p we have

---

[27] We assumed that $a < c < b$, but for $a = c$ and $c = b$, equality (5) is evident because it is clear from the very definition of $T_*(f)$ that $\overset{c}{\underset{c}{\mathbf{T}}}_*(f) = 0$.

$$\overset{b}{\underset{p}{\mathbf{T}}}_*(f) = \overset{b}{\underset{p}{\mathbf{T}}}(f).$$

In this case,

$$\overset{b}{\underset{p}{\mathbf{T}}}_*(f) = \lim \overset{x}{\underset{p}{\mathbf{T}}}(f), \tag{7}$$

where $p < x < b$, $x \to b$. But since

$$\overset{x}{\underset{p}{\mathbf{T}}}(f) = \overset{x}{\underset{p}{\mathbf{T}}}_*(f),$$

instead of (7) we can write

$$\overset{b}{\underset{p}{\mathbf{T}}}_*(f) = \lim \overset{x}{\underset{p}{\mathbf{T}}}_*(f),$$

and it remains to note that

$$\overset{b}{\underset{a}{\mathbf{T}}}_*(f) - \overset{x}{\underset{a}{\mathbf{T}}}_*(f) = \overset{b}{\underset{p}{\mathbf{T}}}_*(f) - \overset{x}{\underset{p}{\mathbf{T}}}_*(f).$$

If b is an isolated point of $S_T(f; [a, b])$, then for p sufficiently close to b, it will be the only T-singular point on the segment $[p, b]$. By the definition of $T_*(f)$, (7) will again be satisfied and the proof is concluded as in the preceding case.

Finally, we consider the fundamental case when b is a limit point of the set $S_T(f; [a, b])$. In this case, the point b cannot serve as a right endpoint of any of the intervals (a_n, b_n), complementary to $S_T(f; [a, b])$, and the number of these intervals will be infinite, as is known. Having noted this, we take an $\varepsilon > 0$ and choose an N such that

$$\sum_{n>N} W_n < \varepsilon. \tag{8}$$

The existence of such an N follows from condition 3), which $f(x)$ must satisfy.

We also select a $\delta > 0$ such that the inequality $b - \delta < x < b$ implies the inequality

$$\left| (L) \int_{S_T([x, b])} f(t)\, dt \right| < \varepsilon. \tag{9}$$

We denote the point $b - \delta, b_1, b_2, \ldots, b_N$, which is farthest to the right, by β and let $x > \beta$.

If $x \in S_T([x, b])$, then by the very definition of $T_*(f)$ we have

$$\overset{b}{\underset{x}{\mathbf{T}}}_*(f) = \sum_{n \in M} I_n + (L) \int_{S_T([x, b])} f(t)\, dt, \tag{10}$$

where M is the set of those n for which $(a_n, b_n) \subset [x, b]$. Clearly, $n > N$ for all these n and hence

$$\left| \sum_{n \in M} I_n \right| \leqslant \sum_{n \in M} W_n < \varepsilon. \tag{11}$$

But if $x \bar{\in} S_T([x, b])$, then we can find an m such that $a_m \leqslant x < b_m$. Clearly, $m > N$. Then, instead of (10), we have

$$\overset{b}{\underset{x}{\mathbf{T}}}_*(f) = \overset{b_m}{\underset{x}{\mathbf{T}}}_*(f) + \sum_{n \in M} I_n + (L) \int_{S_T([x, b])} . \tag{12}$$

But

$$\overset{b_m}{\mathbf{T}}_{*}(f) = \lim \overset{y}{\mathbf{T}}(f) \qquad (x < y < b_m, y \to b_m)$$

and hence

$$|\overset{b_m}{\mathbf{T}}_{*}(f)| \leqslant W_m < \varepsilon.$$

From this, in combination with (9) and (11), it is clear that for $x > \beta$ we have

$$|\overset{b}{\mathbf{T}}_{*}(f) - \overset{x}{\mathbf{T}}_{*}(f)| = |\overset{b}{\mathbf{T}}_{*}(f)| < 3\varepsilon.$$

Thus (6) is valid in this case also and, consequently, $\mathbf{T}_{*}(f)$ is in fact an integral.[28]

§7. THE DENJOY INTEGRAL IN THE RESTRICTED SENSE (I.E. THE DENJOY-PERRON INTEGRAL)

Now, finally, we can give the definition of the Denjoy integral "in the restricted sense." To this end, we shall first construct a transfinite sequence (of the type $\Omega + 1$) of more and more general integrals which we shall call D_ξ-integrals, and which we shall denote by

$$(D_\xi) \int_a^b f(x)\,dx \tag{1}$$

for a function $f(x)$ defined on the closed integral $[a, b]$.

The integrals (1) are defined inductively. Namely,

$$(D_0) \int_a^b f(x)\,dx$$

is to be understood as simply the Lebesgue integral

$$(L) \int_a^b f(x)\,dx.$$

We shall now assume that we have already defined integrals (1) for all $\xi < \eta$, where[29] $\eta \leqslant \Omega$. If η is a number of the first kind and $\eta - 1$ is its immediate predecessor, then, setting

$$\overset{b}{\mathbf{T}}(f) = (D_{\eta-1}) \int_a^b f(x)\,dx,$$

we define the integral D_η by means of the formula

$$(D_\eta) \int_a^b f(x)\,dx = \overset{b}{\mathbf{T}}_{*}(f).$$

But if η is a number of the second kind, then we include in the class of functions

[28] Which is more general than $T(f)$.

[29] We shall explain further on why integrals (1) are not considered for $\xi > \Omega$

which are D_η-integrable on the closed interval $[a, b]$ all functions $f(x)$ which are D_ξ-integrable on this segment for at least one $\xi < \eta$, and we set, by definition,

$$(D_\eta)\int_a^b f(x)\,dx = (D_{\xi_0})\int_a^b f(x)\,dx,$$

where ξ_0 is the smallest[30] of these ξ. Thus, if η is a number of the second kind, then the class of functions which are D_η-integrable is the set-theoretic sum (over all $\xi < \eta$) of classes of D_ξ-integrable functions. Therefore the definition of the integrals (1) which we have just given can be written in terms of the following symbolic relations:

$$D_0 = L,\quad D_\eta = (D_{\eta-1})_*,\quad D_\eta = \sum_{\xi<\eta} D_\xi,$$

where the second relation must be used if η is a number of the first kind and the third if η is of the second kind.

In particular, taking $\xi = \Omega$, we arrive at the Denjoy integral "in the restricted sense."

DEFINITION. The *Denjoy integral* in the restricted sense of the function $f(x)$, defined on the closed interval $[a, b]$, is the integral

$$(D_\Omega)\int_a^b f(x)\,dx. \tag{2}$$

As we have already mentioned, this integral is frequently called the Denjoy-Perron integral. Instead of the notation (2), it is usually denoted by[31]

$$(D)\int_a^b f(x)\,dx.$$

Using the same symbolism as above, we can write the formula

$$D = \sum_{\xi<\Omega} D_\xi,$$

so that every function which is Denjoy-integrable is also D_ξ-integrable for some $\xi < \Omega$ and if the smallest of these ξ is ξ_0, then

$$(D)\int_a^b f(x)\,dx = (D_{\xi_0})\int_a^b f(x)\,dx.$$

We shall stop now to consider certain properties of the integral introduced above.

THEOREM 1. *If the function $f(x)$, defined on the closed interval $[a, b]$, is D_ξ-integrable for some $\xi < \Omega$, then it is also D_η-integrable on $[a, b]$ for all η which satisfy the inequalities $\xi \leqslant \eta \leqslant \Omega$ and*

$$(D_\eta)\int_a^b f(x)\,dx = (D_\xi)\int_a^b f(x)\,dx.$$

---

[30] It is easily seen that this ξ_0 is necessarily a number of the first kind. It is clear that the functional $(D_\eta)\int_a^b f(x)\,dx$ is an integral in the sense of §6.

[31] And sometimes by $(D_*)\int_a^b f(x)\,dx$.

Proof. We shall prove the theorem first for $\xi = 0$. When $\eta = 0$, the assertion is trivial. We shall assume that it is already proved for all $\eta < \zeta$, where $0 < \zeta \leqslant \Omega$. If ζ is a number of the first kind, then the assertion is true for $\eta = \zeta$ inasmuch as the integral T_* is more general than T. But if ζ is a number of the second kind, then the situation reduces immediately to the definition of integral (1), when the index is a number of the second kind. Thus, for $\xi = 0$ the theorem is proved. Suppose it is proved for all $\xi < \alpha$ where $0 < \alpha < \Omega$. We shall verify that it is then also true for $\xi = \alpha$.

The validity of the assertion of the theorem for $\eta = \alpha$ is trivial. We shall assume that it is true for all $\eta < \zeta$, where $\alpha < \zeta \leqslant \Omega$. If ζ is a number of the first kind, then the validity of the assertion for $\eta = \zeta$ is established as above. But suppose ζ is of the second kind. Then the D_ζ-integrability of the function is clear. In this connection,

$$(D_\zeta) \int_a^b f(x)\,dx = (D_\beta) \int_a^b f(x)\,dx,$$

where β is the smallest of the γ's which possess the property that the function $f(x)$ is D_γ-integrable. It is clear that $\beta \leqslant \alpha$. If $\beta = \alpha$, then the assertion is proved; but if $\beta < \alpha$, then it follows from the inductive assumption that the theorem is valid for all $\xi < \alpha$, which is what we wanted to prove.

The theorem we have just proved can be formulated briefly as follows: the larger we take ξ, the more general the integral D_ξ.

We shall agree to denote the set of D_ξ-singular points of the function $f(x)$, defined on $[a, b]$, by $S_\xi(f; [a, b])$. Sometimes we shall use one of the following simpler notations:

$$S_\xi([a, b]),\ S_\xi(f),\ S_\xi.$$

It follows directly from Theorem 1 that $S_\xi \supset S_\eta$ for $\xi < \eta$.

THEOREM 2. *A necessary and sufficient condition that the function $f(x)$, defined on $[a, b]$, be D-integrable on $[a, b]$ is that*

$$\prod_{\xi < \Omega} S_\xi(f; [a, b]) = 0. \tag{3}$$

Proof. Suppose $f(x)$ is D-integrable on $[a, b]$. Then it is D_ξ-integrable on $[a, b]$ for some $\xi < \Omega$ and for this ξ we have $S_\xi(f; [a, b]) = 0$. This proves the necessity of condition (3). The sufficiency of this condition follows from the Cantor-Baire stationary principle[32] inasmuch as when (3) is satisfied, a ξ can be found such that $S_\xi(f; [a, b]) = 0$ and this implies the D_ξ-integrability of $f(x)$ on $[a, b]$.

In concluding this section, we shall stop to consider why one does not consider integrals (1) for $\xi > \Omega$. The crux of the matter is that if we wished to introduce the integral $D_{\Omega+1}$ according to the formula

$$D_{\Omega+1} = (D_\Omega)_*$$

then this would not lead to an extension of the class of integrable functions. In fact, we shall assume that the function $f(x)$, defined on $[a, b]$, is $D_{\Omega+1}$-integrable on this segment but that it is not D_Ω-integrable. In this case, the set $S_\Omega(f)$ would be nowhere dense on $[a, b]$ and $f(x)$ would be summable on it. We shall denote the set of all intervals

[32] Chapt. XV, §3.

complementary to the set $S_\Omega(f)$ by $\{(a_n, b_n)\}$. For each of them, the finite limit of the integral

$$(D_\Omega)\int_\alpha^\beta f(x)\,dx \tag{4}$$

must exist for $a_n < \alpha < \beta < b_n$, $\alpha \to a_n$, $\beta \to b_n$. We fix the number n and let

$$\alpha_1 < \beta_1,\ \alpha_1 > \alpha_2 > \alpha_3 > \ldots,\ \beta_1 < \beta_2 < \beta_3 < \ldots,\ \alpha_k \to a_n,\ \beta_k \to b_n.$$

Then on each of the segments $[\alpha_k, \beta_k]$, our $f(x)$ must be D_Ω-integrable and this means the existence for every natural k of a number $\xi_k < \Omega$ such that $f(x)$ will be D_{ξ_k}-integrable on $[\alpha_k, \beta_k]$. Inasmuch as there are only a denumerable number of numbers ξ_k altogether, we should be able to find a number $\eta_n < \Omega$ which is larger than all the ξ_k. For the same reason, a $\zeta < \Omega$ can be found which will be larger than all the η_n. It is clear that $f(x)$ will be D_ζ-integrable on every segment $[\alpha, \beta]$ which does not intersect $S_\Omega(f)$ and lies in $[a, b]$.

From this it follows that the sets $S_\zeta(f)$ and $S_\Omega(f)$ coincide and that it is possible to replace (D_Ω) in integrals (4) by (D_ζ). All this, taken together, simply means the $D_{\zeta+1}$-integrability of $f(x)$. The latter, however, contradicts the assumption that $f(x)$ is not D_Ω-integrable.

§8. H. HAKE'S THEOREM

In 1921, the German mathematician H. Hake proved that the Perron integral is more general[33] than the Denjoy-Perron integral. We shall discuss this result in this section.

LEMMA 1. *Suppose the function $f(x)$, defined on the segment $[a, b]$, is (P)-integrable on every segment $[a, \beta]$ where $a < \beta < b$. If the finite limit*

$$J = \lim_{\beta \to b} (P)\int_a^\beta f(x)\,dx$$

exists, then $f(x)$ will also be (P)-integrable on all of $[a, b]$, where

$$(P)\int_a^b f(x)\,dx = J. \tag{1}$$

Proof. We set $b_0 = a$ and

$$b_n = \frac{b_{n-1} + b}{2} \qquad (n = 1, 2, 3, \ldots).$$

Clearly, $b_0 < b_1 < b_2 < \ldots$ and $b_n \to b$. The function $f(x)$ will be (P)-integrable on every segment $[b_{n-1}, b_n]$. We choose some $\varepsilon > 0$. For every natural n, there will exist a superfunction $U_n(x)$ for the function $f(x)$ on the segment $[b_{n-1}, b_n]$ such that

[33] We use this term in exactly the sense given it in §6. Thus, every integral T is "more general" than itself. In particular, we encounter this situation in the case under consideration because, as will be shown below, the integrals P and D are identical.

$$U_n(b_n) < (P)\int_{b_{n-1}}^{b_n} f(x)\,dx + \frac{\varepsilon}{2^n}.$$

With the aid of these superfunctions, we define a function $M(x), a \leqslant x < b$, by setting

$M(x) = U_1(x)$ for $a \leqslant x \leqslant b_1$

$M(x) = U_1(b_1) + U_2(x)$ for $b_1 \leqslant x \leqslant b_2$

. . . . . .

$M(x) = U_1(b_1) + U_2(b_2) + \ldots + U_{n-1}(b_{n-1}) + U_n(x)$ for $b_{n-1} \leqslant x \leqslant b_n$

. . . . . .

Since $U_n(b_{n-1}) = 0$, $M(x)$ is continuous and

$$\underline{D}M(x) > -\infty, \quad \underline{D}M(x) \geqslant f(x) \tag{2}$$

for arbitrary x in $[a, b]$.

We shall verify the existence of the finite limit

$$\lim_{x \to b} M(x). \tag{3}$$

By the very definition of $U_n(x)$, we have

$$0 \leqslant U_n(b_n) - (P)\int_{b_{n-1}}^{b_n} f(x)\,dx < \frac{\varepsilon}{2^n} \qquad (n = 1, 2, 3, \ldots)$$

and hence the series

$$\sum_{n=1}^{\infty}\left[U_n(b_n) - (P)\int_{b_{n-1}}^{b_n} f(x)\,dx\right]$$

converges and its sum is less than ε. On the other hand, the series

$$\sum_{n=1}^{\infty}(P)\int_{b_{n-1}}^{b_n} f(x)\,dx$$

converges and its sum equals J. This implies the convergence of the series

$$\sum_{n=1}^{\infty} U_n(b_n), \tag{4}$$

whose sum is less than $J + \varepsilon$.

Now suppose that $b_{n-1} \leqslant x \leqslant b_n$. Then

$$(P)\int_{b_{n-1}}^{x} f(t)\,dt \leqslant U_n(x) < (P)\int_{b_{n-1}}^{x} f(t)\,dt + \frac{\varepsilon}{2^n}. \tag{5}$$

Now $n \to \infty$ when $x \to b$, and the left and right members in (5) tend to zero. Hence we also have that $\lim U_n(x) = 0$. This implies the existence of the limit (3) and this limit is equal to the sum of series (4). Denoting this limit by $M(b)$, we arrive at the function $M(x)$ which is defined and continuous on all of $[a, b]$. However, we still

cannot call it a superfunction for $f(x)$ because inequalities (2) do not necessarily have to be satisfied for $x = b$.[34]

We shall show that starting with $M(x)$, one can construct a function $U(x)$ which will serve as a superfunction for $f(x)$ on all of $[a, b]$. To this end, we denote by $S(x)$ the oscillation of the function $M(x)$ on the segment $[x, b]$. It is clear that $S(x)$ is a decreasing, non-negative function, defined on all of $[a, b]$, where $S(b) = 0$. It is also not difficult to verify that $S(x)$ is continuous on $[a, b]$.

We choose a point c $(a < c < b)$ so close to b that $S(c) < \varepsilon$ and we introduce a function $U(x)$ by setting

$$U(x) = \begin{cases} M(x) \text{ for } a \leqslant x \leqslant c \\ M(x) + S(c) - S(x) + \varepsilon \dfrac{\sqrt{b-c} - \sqrt{b-x}}{\sqrt{b-c}} \text{ for } c \leqslant x \leqslant b. \end{cases}$$

It is clear that $U(x)$ is continuous on $[a, b]$, $U(a) = 0$, and for $a \leqslant x < c$ we have

$$\underline{D}U(x) > -\infty, \quad \underline{D}U(x) \geqslant f(x). \tag{6}$$

These inequalities will also be satisfied for $x = c$ if $\underline{D}U(x)$ is understood to mean the left-sided lower derived number. We shall show that relations (6) will also be satisfied for $x \in [c, b]$. Suppose x and $x + \Delta x$ are two distinct points in $[c, b]$. Then

$$\frac{U(x+\Delta x) - U(x)}{\Delta x} = \frac{M(x+\Delta x) - M(x)}{\Delta x} - \frac{S(x+\Delta x) - S(x)}{\Delta x} -$$

$$- \frac{\varepsilon}{\sqrt{b-c}} \cdot \frac{\sqrt{b-x-\Delta x} - \sqrt{b-x}}{\Delta x}.$$

Since both of the functions $S(x)$ and $\sqrt{b-x}$ decrease, we have

$$\frac{U(x+\Delta x) - U(x)}{\Delta x} \geqslant \frac{M(x+\Delta x) - M(x)}{\Delta x}$$

and relations (6) are valid for $x \neq b$ in virtue of (2).

On the other hand, if $c \leqslant b - h < b$, then

$$\frac{U(b-h) - U(b)}{-h} = \frac{M(b-h) - M(b) - S(b-h)}{-h} + \frac{\varepsilon}{\sqrt{b-c}} \cdot \frac{1}{\sqrt{h}}.$$

According to the definition of $S(x)$, we have $M(b-h) - M(b) \leqslant S(b-h)$ and hence

$$\frac{U(b-h) - U(b)}{-h} \geqslant \frac{\varepsilon}{\sqrt{b-c}} \cdot \frac{1}{\sqrt{h}}.$$

This implies that $U'(b) = +\infty$, i.e. inequalities (6) are also satisfied for $x = b$. Thus, $U(x)$ is a superfunction for $f(x)$. In this connection,

$$U(b) = M(b) + S(c) + \varepsilon$$

and hence $U(b) < J + 3\varepsilon$.

One can construct a subfunction V for the function $f(x)$ in an analogous manner

[34] Of course, it is only essential that the possibility of the equality $\underline{D}M(b) = -\infty$ is not excluded because the function $f(x)$ can be modified at the point $x = b$ without affecting its P-integrability.

such that $V(b) > J - 3\varepsilon$. Inasmuch as ε is arbitrary, it follows from the above discussion that $f(x)$ is P-integrable on $[a, b]$. As far as equality (1) is concerned, it is a consequence of the continuity of the indefinite Perron integral.

LEMMA 2. *If* 1) $f(x)$ *is defined on* $[a, b]$ *and* (P)*-integrable on every* $[\alpha, b]$ *where* $a < \alpha < b$, 2) *the finite limit*

$$\lim_{\alpha \to a} (P) \int_{\alpha}^{b} f(x)\, dx$$

exists, then $f(x)$ *is* (P)*-integrable on all of* $[a, b]$.

Proof. Rather than $f(x)$, we shall consider the function $f^*(x)$, defined on $[-b, -a]$ by means of the formula $f^*(x) = f(-x)$. If $a < \alpha < b$ and $U(x)$ is a superfunction for $f(x)$ on the segment $[\alpha, b]$, then the function

$$U^*(x) = U(b) - U(-x)$$

will be a superfunction for $f^*(x)$ on $[-b, -\alpha]$, because

$$\frac{U^*(x+h) - U^*(x)}{h} = \frac{U(-x-h) - U(-x)}{-h}$$

and hence

$$\underline{D}U^*(x) = \underline{D}U(-x).$$

Since $U^*(-\alpha) = U(b)$, we have $\inf\{U^*(-\alpha)\} = \inf\{U(b)\}$. Using this fact, it is easily verified that $f^*(x)$ satisfies all the conditions of Lemma 1 on the segment $[-b, -a]$. This means that $f^*(x)$ is (P)-integrable on $[-b, -a]$, and this is equivalent to the (P)-integrability of $f(x)$ on $[a, b]$.

Now the two preceding lemmas imply the following theorem.

THEOREM 1. *If* $f(x)$ *is defined on* $[a, b]$ *and is* (P)*-integrable on every* $[\alpha, \beta] \subset (a, b)$ *and the finite limit*

$$\lim (P) \int_{\alpha}^{\beta} f(x)\, dx \qquad (\alpha \to a + 0, \beta \to b - 0)$$

exists, then $f(x)$ *is* (P)*-integrable on all of* $[a, b]$.

For the sequel, we shall need the following lemma.

LEMMA 3. *Suppose* $[a, b]$ *contains a denumerable set of segments* $[a_k, b_k]$ *which are pairwise without interior points. Suppose the function* $f(x)$, *defined on* $[a, b]$, *equals zero in the exterior of the intervals* (a_k, b_k) *and that it is* (P)*-integrable on every* $[a_k, b_k]$. *If*

$$W_k = \sup \left| (P) \int_{a}^{\beta} f(x)\, dx \right| \qquad (a_k \leqslant \alpha \leqslant \beta \leqslant b_k)$$

and

$$\sum_{k=1}^{\infty} W_k = H < +\infty,$$

then, for arbitrary $\varepsilon > 0$ *there exists a superfunction* $U(x)$ *for* $f(x)$ *such that*

$$U(b) \leqslant 3H + \varepsilon. \tag{7}$$

Proof. Let $M_k(x)$ be a superfunction for $f(x)$ on the segment $[a_k, b_k]$ such that

$$M_k(b_k) < (P)\int_{a_k}^{b_k} f(x)\,dx + \frac{\varepsilon}{3\cdot 2^k}.$$

Since

$$(P)\int_{a_k}^{x} f(t)\,dt \leqslant M_k(x) \leqslant (P)\int_{a_k}^{x} f(t)\,dt + \frac{\varepsilon}{3\cdot 2^k}$$

for all x in $[a_{,k} b_k]$, we have

$$-W_k \leqslant M_k(x) \leqslant W_k + \frac{\varepsilon}{3\cdot 2^k}.$$

We denote the oscillation of $M(t)$ on the segments $[a_k, b_k]$, $[a_k, x]$ and $[x, b_k]$, respectively, by R_k, $P_k(x)$, and $Q_k(x)$. Then

$$0 \leqslant P_k(x) \leqslant R_k, \quad 0 \leqslant Q_k(x) \leqslant R_k,$$

$P_k(a_k) = 0$, $Q_k(b_k) = 0$, $P_k(x)$ and $Q_k(x)$ are continuous on $[a_k, b_k]$ where the first of these functions increases and the second decreases.

We shall show that $R_k \leqslant W_k + \frac{\varepsilon}{3\cdot 2^k}$. In fact, if x and y are two points in $[a_k, b_k]$, where $x \leqslant y$, then

$$(P)\int_{a_k}^{x} f(t)\,dt + \frac{\varepsilon}{3\cdot 2^k} \geqslant M_k(x) \geqslant (P)\int_{a_k}^{x} f(t)\,dt,$$

$$(P)\int_{a_k}^{y} f(t)\,dt \leqslant M_k(y) \leqslant (P)\int_{a_k}^{y} f(t)\,dt + \frac{\varepsilon}{3\cdot 2^k},$$

from which it follows that

$$(P)\int_{x}^{y} f(t)\,dt - \frac{\varepsilon}{3\cdot 2^k} \leqslant M_k(y) - M_k(x) \leqslant (P)\int_{x}^{y} f(t)\,dt + \frac{\varepsilon}{3\cdot 2^k}$$

and hence

$$|M_k(y) - M_k(x)| \leqslant W_k + \frac{\varepsilon}{3\cdot 2^k}$$

which proves the required estimate of R_k.

We now define a sequence of functions $\varphi_k(x)$ on all of $[a, b]$ by setting

$$\varphi_k(x) = \begin{cases} 0 \text{ for } a \leqslant x \leqslant a_k, \\ M_k(x) + R_k + P_k(x) - Q_k(x) \text{ for } a_k \leqslant x \leqslant b_k, \\ M_k(b_k) + 2R_k \text{ for } b_k \leqslant x \leqslant b. \end{cases}$$

All these functions are continuous and

$$0 \leqslant \varphi_k(x) \leqslant 3W_k + \frac{\varepsilon}{2^k}.$$

We shall show that

$$U(x)=\sum_{k=1}^{\infty}\varphi_k(x)$$

is the function required. The equality $U(a)=0$, inequality (7), and the continuity of $U(x)$ are obvious. It remains to verify that

$$\underline{D}U(x)>-\infty,\ \underline{D}U(x)\geqslant f(x).$$

We shall prove these inequalities only for the right-sided derived numbers[35] $\underline{D}_+U(x)$, because the reasoning is almost analogous (although somewhat more complicated) for the left-sided derived numbers $\underline{D}_-U(x)$.

Suppose $a\leqslant x_0<b$. If an i can be found such that $a_i\leqslant x_0<b_i$, then $a_i<x_0+h<b_i$ for sufficiently small $h>0$ and hence for $k\neq i$ we have $\varphi_k(x_0+h)=\varphi_k(x_0)$ so that

$$U(x_0+h)-U(x_0)=\varphi_i(x_0+h)-\varphi_i(x_0)\geqslant M_i(x_0+h)-M_i(x_0).$$

This implies both of the following inequalities

$$\underline{D}_+U(x_0)>-\infty,\ \underline{D}_+U(x_0)\geqslant f(x_0). \tag{8}$$

We shall now assume that such an i does not exist. Then x_0 lies in the exterior of all the intervals $[a_k,b_k)$ and hence $f(x_0)=0$. We choose any natural k. If $b_k\leqslant x_0$, then for arbitrary $h>0$ we have $\varphi_k(x_0+h)=\varphi_k(x_0)$. But if $b_k\geqslant x_0$, then also $a_k>x_0$ (because otherwise we should have $a_k\leqslant x_0<b_k$). Hence $\varphi_k(x_0)=0\leqslant\varphi_k(x_0+h)$. Thus, for $h>0$ and arbitrary k we have $\varphi_k(x_0)\leqslant\varphi_k(x_0+h)$. But then $\underline{D}_+U(x_0)\geqslant 0$ and (8) is again satisfied. This completes the proof of the lemma.

We shall now prove a proposition which will be of fundamental importance in the sequel.

THEOREM 2. *If the integral* $\overset{b}{\underset{a}{\mathbf{T}}}(f)$ *is less general*[36] *than the Perron integral, then its extension* $\overset{b}{\underset{a}{\mathbf{T}_*}}(f)$ *is also less general than the Perron integral.*

Proof. Suppose $f(x)$ is a T_*-integrable function on $[a,b]$ but that it is not T-integrable on this segment. We denote the set of T-singular points of $f(x)$ by S and the intervals complementary to the set S by (a_k,b_k). According to Theorem 1, the function $f(x)$ will be (P)-integrable on each of the segments $[a_k,b_k]$.

We take an $\varepsilon>0$ and find an m such that

$$\sum_{k=m+1}^{\infty}W_k<\varepsilon,$$

where, as above,

$$W_k=\sup\left|\overset{\beta}{\underset{\alpha}{\mathbf{T}}}(f)\right|=\sup\left|(P)\int_{\alpha}^{\beta}f(x)\,dx\right| \qquad (a_k<\alpha<\beta<b_k).$$

We now consider three functions $p(x)$, $q(x)$, $r(x)$, defined as follows:

---

[35]The right-sided derived number is the limit of sequences $\dfrac{U(x_0+h_k)-U(x_0)}{h_k}$ in which $h_k\to 0$ and $h_k>0$. The left-sided derived number is defined analogously, by simply taking $h_k<0$.

[36]This means that the integral $(P)\int_a^b f(x)\,dx$ is more general (in the sense of §6) than $\overset{b}{\underset{a}{\mathbf{T}}}(f)$.

$$p(x)=\begin{cases} f(x) \text{ for } x\in\sum_{k=1}^{m}(a_k,b_k), \\ 0 \quad \text{for } x\in[a,b]-\sum_{k=1}^{m}(a_k,b_k); \end{cases}$$

$$q(x)=\begin{cases} f(x) \text{ for } x\in\sum_{k=m+1}^{\infty}(a_k,b_k), \\ 0 \quad \text{for } x\in[a,b]-\sum_{k=m+1}^{\infty}(a_k,b_k); \end{cases}$$

$$r(x)=\begin{cases} f(x) \text{ for } x\in S, \\ 0 \quad \text{for } x\in[a,b]-S. \end{cases}$$

It is clear that

$$f(x)=p(x)+q(x)+r(x).$$

The function $r(x)$ is summable on $[a,b]$ and hence it is P-integrable on this segment and

$$(P)\int_a^b r(x)\,dx=(L)\int_a^b r(x)\,dx=(L)\int_S f(x)\,dx.$$

The function $p(x)$ is also P-integrable on $[a,b]$. In fact, as we have already noted, the function $f(x)$ is integrable on each of the segments $[a_k,b_k]$. If its value is changed at the two points a_k and b_k by setting $f(a_k)=0$ and $f(b_k)=0$ this will not disturb the P-integrability of $f(x)$ and it will not change the magnitude of its integral. Therefore the function $p(x)$ will be P-integrable on the segments $[a_k,b_k]$ $(k=1,2,\ldots,m)$. But this function equals 0 and is manifestly P-integrable on the set $[a,b]-\sum_{k=1}^{m}(a_k,b_k)$. Thus, $p(x)$ is P-integrable on all of $[a,b]$ and

$$(P)\int_a^b p(x)\,dx=\sum_{k=1}^{m}(P)\int_{a_k}^{b_k} f(x)\,dx=\sum_{k=1}^{m} I_k,$$

where we have set

$$I_k=\lim_{\alpha}^{\beta}\mathbf{T}(f) \qquad (\alpha\to a_k+0,\beta\to b_k-0).$$

In virtue of the P-integrability of $p(x)$ and $r(x)$ (and the values of the P-integrals indicated above), we can find superfunctions $U_p(x)$, $U_r(x)$ for $p(x)$ and $r(x)$ respectively, such that

$$U_p(b)<\sum_{k=1}^{m} I_k+\varepsilon,\ U_r(b)<(L)\int_S f(x)\,dx+\varepsilon.$$

Finally, in virtue of Lemma 3, there exists a superfunction $U_q(x)$ for $q(x)$ such that[37] $U_q(b)<4\varepsilon$.

We set

$$U(x)=U_p(x)+U_q(x)+U_r(x).$$

Then (see Corollary to Lemma 1, §2) $U(x)$ will be a superfunction for $f(x)$ and

[37] Here, $H<\varepsilon$.

$$U(b) < \sum_{k=1}^{m} I_k + (L)\int_S f(x)\,dx + 6\varepsilon.$$

On the other hand,

$$\left|\sum_{k=m+1}^{\infty} I_k\right| \leqslant \sum_{k=m+1}^{\infty} W_k < \varepsilon.$$

Hence

$$U(b) < \sum_{k=1}^{\infty} I_k + (L)\int_S f(x)\,dx + 7\varepsilon,$$

or, what amounts to the same thing,

$$U(b) < \overset{b}{\underset{a}{\mathbf{T}}}_*(f) + 7\varepsilon.$$

The remainder of the proof is clear.

We can now easily prove the theorem which interests us.

THEOREM 3 (H. HAKE). *The Perron integral is more general than the Denjoy-Perron integral.*

Proof. Every D-integrable function is D_ξ-integrable for any $\xi < \Omega$. Therefore it suffices to prove the following proposition: if the function is D_ξ-integrable for $\xi < \Omega$ then it is also P-integrable, and its (P) and (D_ξ) integrals coincide. For $\xi = 0$ this proposition is valid. Suppose it is valid for all $\xi < \eta$ and that $f(x)$ is a D_η-integrable function. If η is a number of the second kind, then the validity of the proposition is clear for this η. But if η is a number of the first kind and $\eta - 1$ is its immediate predecessor then, by assumption, our proposition is valid for $D_{\eta-1}$-integrals and it suffices to refer to the fact that

$$D_\eta = (D_{\eta-1})_*,$$

and then apply Theorem 2. This completes the proof of the theorem.

§9. THE P. S. ALEKSANDROV - H. LOOMAN THEOREM

In this section we shall prove that the Perron and Denjoy integrals are completely identical. In virtue of the Hake theorem, to do this it will suffice to show that the Denjoy integral is more general[38] than the Perron integral. This fact was established independently by the Russian mathematician P. S. Aleksandrov (1924) and the Dutch mathematician H. Looman (1925).

LEMMA 1. *Suppose the function $f(x)$ is (P)-integrable on the segment $[a,b]$ and that $U(x)$ is any one of its superfunctions. We denote the set of all points x in $[a,b]$ such that*

$$\frac{U(t) - U(x)}{t - x} \geqslant -i \tag{1}$$

for arbitrary $t \neq x$ and $t \in [a,b]$ by E_i $(i = 1, 2, 3, \ldots)$. Then a) each of the sets E_i is closed, b) the formula

$$[a,b] = \sum_{i=1}^{\infty} E_i \tag{2}$$

[38]See first footnote to this section.

is valid, c) the function $f(x)$ is summable on each of the sets E_i.

Proof. Suppose the sequence $\{x_k\}$ of points in E_i has the limit point x^*. We choose any $t \in [a, b]$, different from x^*. Then, for sufficiently large k, we have $x_k \neq t$ and

$$\frac{U(t) - U(x_k)}{t - x_k} \geqslant -i.$$

We now use the continuity of $U(x)$ and pass to the limit as $k \to \infty$. This proves *a)*. In order to prove *b)*, we choose any $x_0 \in [a, b]$; by the definition of a superfunction, we have

$$\underline{\lim} \frac{U(t) - U(x_0)}{t - x_0} = \underline{D}U(x_0) > -\infty.$$

We take any number A which satisfies the inequality

$$A < \underline{\lim} \frac{U(t) - U(x_0)}{t - x_0}.$$

By the very definition of the limit inferior, there exists a $\delta > 0$ such that for $t \in [a, b](x_0 - \delta, x_0 + \delta)$ (and $t \neq x_0$) we have

$$\frac{U(t) - U(x_0)}{t - x_0} > A.$$

On the other hand, if $t \in [a, b]$ and $t \bar{\in} (x_0 - \delta, x_0 + \delta)$, then

$$\left| \frac{U(t) - U(x_0)}{t - x_0} \right| \leqslant \frac{2M}{\delta},$$

where we have set $M = \max |U(x)|$. By the same token, we have, for these i, that

$$\frac{U(t) - U(x_0)}{t - x_0} \geqslant -\frac{2M}{\delta}.$$

If i is so large that

$$-i < \min \left\{ A, -\frac{2M}{\delta} \right\},$$

then $x_0 \in E_i$.

It remains to prove *c)*.[39] To this end, we choose any i and set

$$U_i(x) = U(x) + ix.$$

If $t \neq x$, then

$$\frac{U_i(t) - U_i(x)}{t - x} = \frac{U(t) - U(x)}{t - x} + i$$

and hence, for $x \in E_i$ and $t \neq x$, we have

$$\frac{U_i(t) - U_i(x)}{t - x} \geqslant 0.$$

It follows from this that the function $U_i(x)$ increases on the set E_i. By the Cantor-Bendikson theorem, E_i can be written in the form $P_i + D_i$, where P_i is perfect and D_i is at most denumerable. We denote the leftmost and rightmost points of P_i by $\bar{a}$ and $\bar{b}$ respectively and let $\{(a_n, b_n)\}$ be the set of intervals complementary to P_i with respect

[39] Of course, *c)* holds only for $mE_i > 0$. We shall assume that this is the case.

to $[\bar{a}, \bar{b}]$. We define the function $\bar{U}_i(x)$ on $[\bar{a}, \bar{b}]$ by setting it equal to $U_i(x)$ on the set P_i and letting it be linear on each of the segments $[a_n, b_n]$. This function is increasing on $[\bar{a}, \bar{b}]$. Therefore it has a finite derivative $\bar{U}_i'(x)$ almost everywhere on $[\bar{a}, \bar{b}]$ and the latter is summable on $[\bar{a}, \bar{b}]$ and a fortiori on P_i.

On the other hand, the superfunction $U(x)$ has a finite derivative $U'(x)$ almost everywhere on $[a, b]$ (see §4). Since $U_i(x) = U(x) + ix$, we have that $U_i(x)$ is also almost everywhere differentiable on $[a, b]$, and hence also on $[\bar{a}, \bar{b}]$. If x_0 is a point of P_i at which both derivatives $\bar{U}_i'(x_0)$ and $U_i'(x_0)$ exist, then the latter must necessarily be equal because x_0 is a limit point of the set P_i on which $\bar{U}_i(x)$ and $U_i(x)$ coincide. This means that the derivative $U_i'(x)$ is equivalent to the summable function $\bar{U}_i'(x)$ on P_i and hence it is itself summable. But then $U'(x) = U_i'(x) - i$ is also summable on P_i.

We set

$$F(x) = (P) \int_a^x f(t)\, dt.$$

The difference $U(x) - F(x)$ increases on $[a, b]$, and hence it has a finite derivative almost everywhere on $[a, b]$ which is summable on $[a, b]$ and a fortiori on P_i. But

$$F(x) = U(x) - [U(x) - F(x)].$$

It follows from this that $F'(x)$ is summable on P_i and since $F'(x) = f(x)$ almost everywhere on $[a, b]$, we have that $f(x)$ is summable on P_i. Inasmuch as the summability of $f(x)$ on the denumerable set D_i is trivial, this completes the proof of *c*).

LEMMA 2. *If $f(x)$ is P-integrable on $[a, b]$, then there exists a segment $[c, d] \subset [a, b]$ on which $f(x)$ is summable.*

Proof. We apply the theorem which asserts that a closed set is not a set of the first category on itself to the segment $[a, b]$. According to the corollary to this theorem (Chapt. XV, §3) and relation (2), we can find a segment $[c, d] \subset [a, b]$ and a natural number i such that $[c, d] \subset E_i$. Since $f(x)$ is summable on E_i, it is a fortiori summable on $[c, d]$. This is what we were required to prove.

Since the lemma just proved can be applied to an arbitrary segment $[a_1, b_1] \subset [a, b]$, it implies the following lemma.

LEMMA 3. *If the function $f(x)$ is P-integrable on $[a, b]$, then the set S_0 of the points where it is non-summable (i.e. the set of L-singular points) is nowhere dense.*

Using the notation of §7, we have $S_0 \supset S_1 \supset S_2 \supset \ldots \supset S_\xi \supset \ldots$.

Therefore each of the sets S_ξ is a fortiori nowhere dense.

LEMMA 4. *If the function $f(x)$ is P-integrable on $[a, b]$ and $S_\xi = S_\xi(f;\ [a, b]) \neq 0$, then $S_\xi - S_{\xi+1}$ is also $\neq 0$.*

Proof. Let us first suppose that S contains an isolated point c. We shall assume that $a < c < b$ (if $c = a$ or $c = b$, the reasoning is practically the same). Then we can find a p and q ($a \leqslant p < c < q \leqslant b$), such that c will be the only D_ξ-singular point on $[p, q]$. Suppose r and s are such that $p < r < c < s < q$. On each of the segments $[p, r]$ and $[s, q]$ the function $f(x)$ will be D_ξ-integrable and[40]

---

[40] $(P) \int_a^x f(t)\, dt$ is denoted by $F(x)$. Here, we use that part of the Hake theorem which asserts that the value of the *D*-integral equals the value of the *P*-integral.

$$(D_\xi)\int_p^r f(x)\,dx = F(r) - F(p), \qquad (D_\xi)\int_s^q f(x)\,dx = F(q) - F(s).$$

In virtue of the continuity of $F(x)$, the above integrals tend to a finite limit as $r \to c$, $s \to c$. It follows from this that the function $f(x)$ is $D_{\xi+1}$-integrable on $[p, c]$ and on $[c, q]$ and hence also on $[p, q]$. In other words, $c \bar{\in} S_{\xi+1}$, and the theorem is proved.

We shall now consider the case when S_ξ contains no isolated points, i.e. when S_ξ is a perfect set. We consider the function $U(x)$ and sets E_i which we already considered in Lemma 1.

Furthermore, we denote any subfunction for $f(x)$ by $V(x)$ and construct for each natural number j a set F_j of those x for which

$$\frac{V(t) - V(x)}{t - x} \leqslant j$$

for $t \neq x$.

Analogous to the sets E_i, the sets F_j are closed and their sum constitutes the entire segment $[a, b]$. In this case,

$$S_\xi = \sum_{i,j} S_\xi E_i F_j.$$

Therefore (on the basis of the corollary to Theorem 2, §3, Chapt. XV, which we have already mentioned above), we can find an i_0 and a j_0 and an interval (h, g) such that

$$0 \neq (h, g) S_\xi \subset E_{i_0} F_{j_0}.$$

Suppose x_0 is any point in $(h, g)S_\xi$. Since S_ξ is nowhere dense, points p and q can be found such that $h < p < x_0 < q < g$ and $p \bar{\in} S_\xi$, $q \bar{\in} S_\xi$. It is clear that

$$0 \neq [p, q] S_\xi \subset E_{i_0} F_{j_0}.$$

The fact that p and q do not occur in S_ξ implies[41] that

$$[p, q] S_\xi = S_\xi([p, q]).$$

Thus, $S_\xi([p, q]) \subset E_{i_0} F_{j_0}$. We shall show that the function is $D_{\xi+1}$-integrable on the segment $[p, q]$. The validity of the lemma will follow from this since the interior of the segment $[p, q]$ cannot contain any point of $S_{\xi+1}$, whereas, as we know, points of S_ξ do lie in the interior of $[p, q]$ (for example, the point x_0).

The condition that $f(x)$ is summable on the (nowhere dense!) set $S_\xi([p, q])$ is satisfied in virtue of the fact that this set is included in E_{i_0} and by Lemma 1. We denote the set of all intervals complementary to the set $S_\xi([p, q])$ with respect to the segment $[p, q]$ by $\{(a_n, b_n)\}$. If n is fixed and $a_n < \alpha < \beta < b_n$, then $f(x)$ is D_ξ-integrable on $[\alpha, \beta]$, and by the Hake theorem we have

$$(D_\xi)\int_\alpha^\beta f(x)\,dx = F(\beta) - F(\alpha).$$

In virtue of the continuity of the indefinite Perron integral, it follows immediately from this that the limit of the D_ξ-integral, written above, exists and is finite as $\alpha \to a_n + 0$, $\beta \to b_n - 0$.

---

[41] If we had $p \in S_\xi$ then p could be the left endpoint of the interval complementary to S_ξ, and then we hould have $p \in [p, q]\, S_\xi - S_\xi([p, q])$.

Therefore, in order to complete the proof of the lemma, it only remains to verify that

$$\sum_n W_n < +\infty,$$

where, as up to this point,

$$W_n = \sup \left| (D_\xi) \int_\alpha^\beta f(x)\,dx \right| \qquad (a_n < \alpha < \beta < b_n).$$

To this end, we note that for all[42] n we have

$$a_n \in S_\xi([p,q]) \subset E_{i_0} F_{j_0}.$$

This means that

$$\frac{U(t)-U(a_n)}{t-a_n} \geqslant -i_0, \quad \frac{V(t)-V(a_n)}{t-a_n} \leqslant j_0$$

for all t in $[a,b]$ which are distinct from a_n.

In particular, if $a_n < t < b_n$, then

$$U(t)-U(a_n) \geqslant -i_0(t-a_n),\ V(t)-V(a_n) \leqslant j_0(t-a_n)$$

and, a fortiori, for these t we have

$$U(t)-U(a_n) \geqslant -i_0(b_n-a_n),\ V(t)-V(a_n) \leqslant j_0(b_n-a_n).$$

On the other hand, both of the functions

$$L(t) = U(t) - F(t),\ M(t) = F(t) - V(t)$$

increase. Hence

$$F(t)-F(a_n) = [U(t)-U(a_n)] - [L(t)-L(a_n)] \geqslant -i_0(b_n-a_n) - [L(b_n)-L(a_n)],$$
$$F(t)-F(a_n) = [V(t)-V(a_n)] + [M(t)-M(a_n)] \leqslant j_0(b_n-a_n) + [M(b_n)-M(a_n)].$$

It follows that

$$|F(\beta)-F(\alpha)| \leqslant (i_0+j_0)(b_n-a_n) + [L(b_n)-L(a_n)] + [M(b_n)-M(a_n)]$$

and hence that also

$$W_n \leqslant (i_0+j_0)(b_n-a_n) + [L(b_n)-L(a_n)] + [M(b_n)-M(a_n)].$$

But then it is clear that $\sum W_n < +\infty$. This completes the proof of the theorem.

THEOREM (P. S. ALEKSANDROV-H. LOOMAN). *A function which is Perron-integrable is also integrable in the Denjoy-Perron sense.*

In fact, according to the Cantor-Baire stationary principle there exists an $\alpha < \Omega$ such that $\prod_{\xi<\Omega} S_\xi = S_\alpha = S_{\alpha+1}$.

If the set $\prod S_\xi$ were non-void, then we should obtain a contradiction of Lemma 4. It remains to apply Theorem 2, §7.

§10. THE DENJOY INTEGRAL IN THE WIDE SENSE (I. E. THE DENJOY-KHINCHIN INTEGRAL)

In order to give the concept of the Denjoy integral in the wide sense, or the Denjoy-Khinchin integral, we must return to the general considerations of §6.

[42] Furthermore, it is the only n for which $a_n = p$ (such an n exists inasmuch as $p \overline{\in} S_\xi$).

Suppose $\overset{b}{\underset{a}{\mathbf{T}}}(f)$ is an arbitrary given integral. We shall construct a generalization $\overset{b}{\underset{a}{\mathbf{T}^*}}(f)$ of it which is distinct from the generalization $\overset{b}{\underset{a}{\mathbf{T}_*}}(f)$ with which we became acquainted in §6.

Namely, we shall agree to include in the class $T^*([a,b])$ every function $f(x)$ which is defined on $[a,b]$ and satisfies the following four conditions:

1) the set $S_T = S_T(f;[a,b])$ is nowhere dense and $f(x)$ is not Lebesgue-summable on S_T;

2) if $\{(a_n, b_n)\}$ is the set of intervals complementary[43] to S_T, then the limit

$$I_n = \lim \overset{\beta}{\underset{\alpha}{\mathbf{T}}}(f) \qquad (a_n < \alpha < \beta < b_n, \alpha \to a_n, \beta \to b_n)$$

exists and is finite for every n;

3) the inequality

$$\sum_n |I_n| < +\infty \tag{1}$$

holds;

4) if there are an infinite number of intervals (a_n, b_n) and

$$W_n = \sup |\overset{\beta}{\underset{\alpha}{\mathbf{T}}}(f)| \qquad (a_n < \alpha < \beta < b_n),$$

then

$$\lim W_n = 0. \tag{2}$$

We see that the difference between the definition of the class $T^*([a,b])$ and the definition of the class $T_*([a,b])$ consists in replacing the single requirement

$$\sum_n W_n < +\infty \tag{3}$$

by the two requirements (1) and (2). Since $|I_n| \leqslant W_n$, (1) and (2) follow from (3) from which we have

$$T_*([a,b]) \subset T^*([a,b]).$$

It follows from this, in particular, that the classes $T^*([a,b])$ are non-void. Without modification of the line of reasoning in §6, we can show that the system of these classes is legitimate.

Having introduced the classes $T^*([a,b])$, we can define a functional $\overset{b}{\underset{a}{\mathbf{T}^*}}$ on each of them by associating the number

$$\overset{b}{\underset{a}{\mathbf{T}^*}}(f) = \sum_n I_n + (L)\int_{S_T} f(x)\,dx$$

with each $f \in T^*([a,b])$. Thus, $\overset{b}{\underset{a}{\mathbf{T}^*}}(f)$ is given by the same formula as $\overset{b}{\underset{a}{\mathbf{T}_*}}(f)$, but on a more extensive class of functions. Following the same line of reasoning as in §6, we can show that $\overset{b}{\underset{a}{\mathbf{T}^*}}(f)$ is an integral in the sense of §6. It is clear from what we have just stated that this integral is more general than $\overset{b}{\underset{a}{\mathbf{T}_*}}(f)$.

---

[43] Under the same assumptions as were made in §6 relative to the leftmost and rightmost of the intervals (a_n, b_n).

The Denjoy integral in the wide sense is constructed with the aid of the integral T^* in exactly the same way that the integral in the restricted sense was constructed with the aid of T_*.

Namely, we introduce the transfinite (of type $\Omega+1$) sequence of integrals

$$(D_\xi)\int_a^b f(x)\,dx, \tag{4}$$

defined by induction. For $\xi=0$, integral (4) is simply the Lebesgue integral. If integrals (4) have been defined for all $\xi<\eta$, where $\eta\leqslant\Omega$, then we set

$$D^\eta=(D^{\eta-1})^*$$

for all η of the first kind and we set

$$D^\eta=\sum_{\xi<\eta}D^\xi$$

for η of the second kind.

The integral

$$D^\Omega=\sum_{\xi<\Omega}D^\xi$$

is the Denjoy integral in the wide sense.

It can be proved that the Denjoy integral in the wide sense is more general than the Denjoy integral in the restricted sense. Moreover, we have[44] in general that the integral D^ξ is more general than D_ξ for arbitrary $\xi\leqslant\Omega$.

EXERCISES FOR CHAPTER XVI

1. The concept of Perron integral is not extended if in the definitions of the superfunction $U(x)$ and subfunction $V(x)$ of the function $f(x)$ we require the satisfaction of the inequalities $\underline{D}U(x)\geqslant f(x)$ and $\bar{D}V(x)\leqslant f(x)$ only almost everywhere.

2. Let S be a closed nowhere dense set on $[a,b]$, let $\{(a_n,b_n)\}$ be the set of its complementary intervals, let T be some integral, and let $f(x)$ be a function defined on $[a,b]$ and T-integrable on every $[\alpha,\beta]\subset(a_n,b_n)$ $(n=1,2,3,\ldots)$. Let

$$W_n=\sup|\mathbf{T}_\alpha^\beta(f)| \qquad (a_n<\alpha<\beta<b_n).$$

If: *a*) the limit

$$I_n=\lim\mathbf{T}_\alpha^\beta(f) \qquad (\alpha\to a_n+0,\ \beta\to b_n-0)$$

exists and is finite for every n and *b*) $\sum W_n<+\infty$, then to every $\varepsilon>0$ there corresponds a $\delta>0$ such that for an arbitrary finite or denumerable system of segments $[\alpha_i,\beta_i]$ for which $[\alpha_i,\beta_i]\subset(a_{n_i},b_{n_i})$ $(n_i\neq n_k$ for $i\neq k)$ and $\sum(\beta_i-\alpha_i)>\sum(b_n-a_n)-\delta$, we have

$$|\sum_n I_n-\sum_i\mathbf{T}_{\alpha_i}^{\beta_i}(f)|<\varepsilon.$$

3. With the same notation as above, let the number H have the property that to

[44] See I. P. Natanson and G. I. Natanson, *On the reciprocal relationship between the Denjoy integrals in the restricted and in the wide sense*, Uspekhi Matem. Nauk, 1958, 13, No. 1.

every $\varepsilon > 0$ there corresponds a $\delta > 0$ such that for an arbitrary finite system of segments $[\alpha_i, \beta_i]$ for which $[\alpha_i, \beta_i] \subset (a_{n_i}, b_{n_i})$ $(n_i \neq n_k$ for $i \neq k)$ and $\sum(\beta_i - \alpha_i) > \sum(b_n - a_n) - \delta$, we have

$$|H - \sum_i \overset{\beta_i}{\underset{\alpha_i}{\mathbf{T}}}(f)| < \varepsilon.$$

Then *a*) and *b*) in Exercise 2 hold.

4. If S_ξ is the set of D_ξ-singular points of the function $f(x)$, then

$$\prod_{\xi < \Omega} S_\xi = S_\Omega.$$

5. A necessary and sufficient condition that the function $f(x)$ be D-integrable is that the relation $S_\xi \neq 0$ imply the relation $S_\xi \neq S_{\xi+1}$.

6. Show that the function $U(x)$ constructed in Lemma 3, §8, satisfies the relations $-\infty \neq \underline{D}_- U(x) \geqslant f(x)$.

7. Prove that if $\overset{b}{\underset{a}{\mathbf{T}}}(f)$ is an integral then $\overset{b}{\underset{a}{\mathbf{T}}}{}^*(f)$ is also an integral.

8. If $f(x) \in D_\xi([a, b])\, D^\eta([a, b])$, then its D_ξ-integral and D^η-integral are equal.

9. The integral D^ξ is more general than D_ξ.

10. For arbitrary $\xi < \Omega$, there exists an $f(x)$ which lies in $D_{\xi+1}([a, b]) - D^\xi([a, b])$.

11. For arbitrary $\xi < \Omega$ we have

$$D^{\xi+1}([a, b]) \neq D_{\xi+1}([a, b]) + D^\xi([a, b]).$$

CHAPTER XVII

SOME IDEAS FROM FUNCTIONAL ANALYSIS

§1. METRIC AND, IN PARTICULAR, NORMED LINEAR SPACES

We have already discussed the concept of a *metric space*, in Chapter VII. Let us recall that a set E of elements $x, y, z, \ldots$ is called a metric space if to an arbitrary pair x, y of elements in E there corresponds a real number $\rho(x, y)$ possessing the following properties:

1) $\rho(x, y) \geqslant 0$, where $\rho(x, y) = 0$ if and only if $x = y$;

2) $\rho(x, y) = \rho(y, x)$;

3) $\rho(x, z) \leqslant \rho(x, y) + \rho(y, z)$.

The number $\rho(x, y)$ is called the *distance* between the elements x and y; inequality 3) is called the *triangle inequality* because in the case of two-dimensional Euclidean space it specifies that the length of one side of a triangle is not greater than the sum of the lengths of the other two sides.

The concept of a *limit of a sequence* of elements $x_1, x_2, x_3, \ldots$ can be introduced in a metric space; this limit is understood to be an element x of the space for which

$$\lim_{n \to \infty} \rho(x_n, x) = 0.$$

The fact that the element x is a limit of the sequence $\{x_n\}$ is written in the usual way:

$$\lim x_n = x, \text{ or } x_n \to x.$$

It is not difficult to show that a sequence of elements in a metric space can have at most one limit.

In fact, if $x_n \to x$ and $x_n \to y$, then

$$0 \leqslant \rho(x, y) \leqslant \rho(x, x_n) + \rho(x_n, y),$$

from which, passing to the limit as $n \to \infty$, we find that

$$\rho(x, y) = 0, \text{ and hence } x = y.$$

DEFINITION. A sequence $\{x_n\}$ of elements in a metric space is said to be *convergent in itself* if to every $\varepsilon > 0$ there exists an N such that for $n > N$ and $m > N$, we have

$$\rho(x_n, x_m) < \varepsilon.$$

It is easy to see that *a sequence having a limit is convergent in itself.* In fact, let $x_n \to x$. Then, taking an arbitrary $\varepsilon > 0$, we can find an N such that for $n > N$, we have $\rho(x_n, x) < \frac{\varepsilon}{2}$. It is clear that when both indices n and m become greater than N, the inequality

$$\rho(x_n, x_m) < \varepsilon$$

is valid.

The converse statement is not true in general. If, for example, we take for our metric space the set of all *nonzero* real numbers, where distance is defined in the usual way, i.e. by the formula $\rho(x,y) = |x-y|$, then the sequence $\left\{\frac{1}{n}\right\}$ will be convergent in itself, but it does not have a limit. It is clear that this situation is due to an unusual "incompleteness" of the space considered, and by "completing" the space by adjoining the element 0, we can remove this defect. In this connection, we introduce the following exact definition.

DEFINITION. A metric space E is said to be *complete* if every sequence of its elements, which is convergent in itself, has a limit in E.

It is well-known that n-dimensional Euclidean spaces possess the property of completeness. The spaces L_p and l_p considered in Chapter VII also possess this property, as was shown.

An important special case of a metric space is a so-called *normed linear space.*

DEFINITION. A set E of elements $x, y, z, \ldots$ is said to be a *linear space* if

I. *To every pair of elements x and y of E there corresponds a third element, $z = x + y$, called the "sum" of x and y.*

II. *To every element $x \in E$ and every real number a, there corresponds an element ax, which is called the "product" of a and x.*

III. *The operations introduced above possess the following properties:*

1) $x + y = y + x$, *i.e. addition is commutative*;
2) $(x + y) + z = x + (y + z)$, *i.e. addition is associative*;
3) *The relation $x + y = x + z$ implies $y = z$*;
4) $1 \cdot x = x$;
5) $a(bx) = (ab)x$;
6) $(a + b)x = ax + bx$;
7) $a(x + y) = ax + ay$.

Let E be a linear space. Correspond to each of its elements x the product $\Theta_x = 0 \cdot x$.

We shall show that this product is indeed independent of the choice of x. In fact, first of all

$$x + \Theta_x = 1 \cdot x + 0 \cdot x = (1 + 0)x = 1 \cdot x = x.$$

But then for an arbitrary pair of elements x and y we have

$$(x + y) + \Theta_x = x + (y + \Theta_x) = x + (\Theta_x + y) = (x + \Theta_x) + y = x + y,$$
$$(x + y) + \Theta_y = x + (y + \Theta_y) = x + y.$$

Thus, $(x + y) + \Theta_x = (x + y) + \Theta_y$, from which it follows that

$$\Theta_x = \Theta_y.$$

Henceforth we shall denote the element Θ_x simply by Θ. It is clear that this element plays the role of zero in the addition of elements in E. In this connection, we shall sometimes denote it by 0 (where, of course, the danger of confusing it with the number 0 must be excluded). It is not difficult to see that besides the element Θ there are no elements z of the space which satisfy the relation $x + z = x$ for arbitrary x. In fact, it would follow from this relation that $x + z = x + \Theta$ and $z = \Theta$.

Thus, a linear space E also possesses the following property:

8) *There exists an element Θ in E such that for all x of E, $x + \Theta = x$. If for at least one $x \in E$ we have $x + z = x$, then $z = \Theta$.*

Further, we note that

9) *If* $ax = \Theta$, *then either* $a = 0$ *or* $x = \Theta$. *Conversely, each of the relations* $a = 0$, $x = \Theta$ *implies the equality* $ax = \Theta$.

In fact, if $a = 0$, then $ax = \Theta$ by the very definition of the element Θ. Further, $0\Theta = \Theta$ and hence, $a\Theta = a\cdot(0\Theta) = (a\cdot 0)\Theta = 0\cdot\Theta = \Theta$. Thus, each of the relations $a = 0$, $x = \Theta$ implies the equality $ax = \Theta$. Now let it be given that $ax = \Theta$. If $a \neq 0$, then $x = 1\cdot(x) = \left(\frac{1}{a}\cdot a\right)x = \frac{1}{a}\cdot(ax) = \frac{1}{a}\Theta = \Theta$. From 9) we deduce without difficulty that

10) *If* $x \neq \Theta$, *then* $ax = bx$ *implies* $a = b$;

11) *If* $a \neq 0$, *then* $ax = ay$ *implies* $x = y$.

In fact, the relation $ax = bx$ shows that $(a-b)x = ax + (-b)x = bx + (-b)x = [b + (-b)]x = 0\cdot x = \Theta$, from which it follows, for $x \neq \Theta$, that $a = b$.

In exactly the same way, from $ax = ay$, we deduce successively that $a[x + (-1)y] = ax + (-a)y = ay + (-a)y = [a + (-a)]y = 0\cdot y = \Theta$. Hence, for $a \neq 0$, we have $x + (-1)y = \Theta$. But, on the other hand, $y + (-1)y = \Theta$, and therefore $x = y$.

We agree to use the following notation henceforth:

$$(-1)\cdot x = -x, \quad x + (-y) = x - y.$$

It is easy to show that

12) $x - x = \Theta$, *and if* $x - y = \Theta$ *then* $x = y$;

13) $x - y = x - z$ *implies* $y = z$;

14) *If* $x + y = z$ *then* $x = z - y$, *and conversely.*

We shall not stop to carry out the simple verification of these statements.

DEFINITION. A linear space E is said to be *normed* if to each of its elements x there corresponds a real number $\|x\|$ called the *norm* of this element and possessing the following properties:

1) $\|x\| \geqslant 0$, *where* $\|x\| = 0$ *if and only if* $x = \Theta$;

2) $\|ax\| = |a| \cdot \|x\|$, *and in particular* $\|-x\| = \|x\|$;

3) $\|x + y\| \leqslant \|x\| + \|y\|$.

It is clear that, having such a space at one's disposal, it is sufficient to set

$$\rho(x, y) = \|x - y\|, \tag{1}$$

in order to transform it into a metric space. Henceforth, when speaking of a normed linear space, we shall always assume it to be metric where the distance is assumed to be given by formula (1).

Complete normed linear spaces were objects of investigation by the Polish mathematician S. S. Banach; for this reason they are also called *Banach spaces.*

The spaces L_p and l_p, considered in Chapter VII, are examples of Banach spaces. We shall now introduce two more examples of such spaces.

Space C. Consider the set of all functions $f(x)$ defined and continuous on the same closed interval $[a, b]$. In this set, we define addition of the elements $f_1 = f_1(x)$ and $f_2 = f_2(x)$ and the multiplication of an element $f = f(x)$ by a number c by means of the usual formulas

$$f_1 + f_2 = f_1(x) + f_2(x), \quad cf = cf(x).$$

It is easy to see that by doing this we transform our set into a linear space. Setting

$$\|f\| = \max |f(x)|$$

for every element $f=f(x)$, we transform our linear space into a normed space. In fact, it can be verified without difficulty that all three norm properties are fulfilled. The linear normed space constructed in this manner is usually denoted by C or by $C([a,b])$.

Let $\{f_n\}$ [where $f_n=f_n(x)$] be a sequence of elements in C. It is clear that the relation $f_n \to f$ [where $f=f(x)\in C$] indicates the uniform convergence of the sequence of the functions $f_n(x)$ to the function $f(x)$.

THEOREM 1. *The space C is complete.*

In fact, let the sequence $\{f_n\}$ $[f_n=f_n(x)]$ of elements in C converge in itself. Fix some point $x\in[a,b]$. Since

$$|f_n(x)-f_m(x)|\leqslant\|f_n-f_m\|,$$

the numerical sequence $\{f_n(x)\}$ converges in itself and (by the well-known Bolzano-Cauchy criterion) the limit

$$\lim_{n\to\infty} f_n(x)=f(x) \tag{2}$$

exists and is finite. Since x is arbitrary, the function $f(x)$ is defined at all points of $[a,b]$. Further, taking $\varepsilon>0$ and finding N such that for $n>N$, $m>N$ we have

$$\|f_n-f_m\|<\varepsilon,$$

we will have, for all x (with $n>N, m>N$), that

$$|f_n(x)-f_m(x)|<\varepsilon.$$

Fix x and $n>N$, and take the limit for increasing m. We thus arrive at the inequality

$$|f_n(x)-f(x)|\leqslant\varepsilon,$$

which is valid for $n\geqslant N$ and arbitrary x. Since the number N is independent of the choice of x, the convergence in (2) is uniform, from which it follows that the limit function $f(x)$ is continuous, and hence it is an element of the space C. Furthermore, the uniform convergence of (2) also means that $f=f(x)$ is the limit function of the sequence $\{f_m\}$ in the sense of the metric of the space C.

Space m. Consider the set of all possible *bounded* sequences[1]

$$x=(x_1,x_2,x_3,\ldots)$$

of real numbers. In this set we define addition, multiplication by a number, and a norm in the following way: let $x=(x_1,x_2,x_3,\ldots)$ and $y=(y_1,y_2,y_3,\ldots)$ be two bounded sequences and let a be some number. Set, by definition,

$$x+y=(x_1+y_1,x_2+y_2,x_3+y_3,\ldots),$$
$$ax=(ax_1,ax_2,ax_3,\ldots),$$
$$\|x\|=\sup\{|x_k|\}.$$

It is not difficult to verify that this transforms our set into a normed linear space. It is customary to denote this space by the letter m.

THEOREM 2. *The space m is complete.*

Let $\{x^{(n)}\}$, where

---

[1]We recall that the numerical sequence $(x_1,x_2,x_3,\ldots)$ is said to be bounded if there exists a constant A such that $|x_k|\leqslant A$ for all k. Thus, the boundedness of a sequence does not at all mean that it has a finite number of terms. On the contrary, there will be a denumerable number of terms (although there may be equal terms).

$$x^{(n)} = (\xi_1^{(n)}, \xi_2^{(n)}, \xi_3^{(n)}, \ldots) \qquad (n = 1, 2, 3, \ldots),$$

be a convergent-in-itself sequence of elements in m.

Since, for arbitrary fixed k, we have

$$|\xi_k^{(n)} - \xi_k^{(m)}| \leqslant \|x^{(n)} - x^{(m)}\|,$$

the numerical sequence $\xi_k^{(1)}, \xi_k^{(2)}, \xi_k^{(3)}, \ldots$ converges in itself and hence it has a finite limit

$$\xi_k = \lim_{n\to\infty} \xi_k^{(n)}.$$

Taking $\varepsilon > 0$, we find an N such that for $n \geqslant N$ and $m \geqslant N$, $\|x^{(n)} - x^{(m)}\| < \varepsilon$. Then for arbitrary k, we have

$$|\xi_k^{(n)} - \xi_k^{(m)}| < \varepsilon.$$

Fixing k and $n \geqslant N$ in the last inequality, we let m tend to infinity. In the limit we obtain the inequality

$$|\xi_k^{(n)} - \xi_k| \leqslant \varepsilon. \tag{3}$$

It follows from this, in particular, that $|\xi_k| \leqslant |\xi_k^{(n)}| + \varepsilon \leqslant \|x^{(n)}\| + \varepsilon$, so that the sequence $x = (\xi_1, \xi_2, \xi_3, \ldots)$ is bounded and hence is included in m. Furthermore, inasmuch as the index k is arbitrary, it follows from (3) that $\|x^{(n)} - x\| \leqslant \varepsilon$, and since to obtain this inequality we needed only that n be $> N$, we have proved the relation

$$\lim_{n\to\infty} x^{(n)} = x.$$

This completes the proof of the theorem.

One must not get the idea that metric spaces other than normed metric spaces do not exist. It is easy to introduce examples of *linear metric spaces which are not normed.*

Space s. Consider the set of all possible numerical sequences $x = (x_1, x_2, x_3, \ldots)$, in which the addition of elements and the multiplication of an element by a number are defined as in the case of the space m.

We agree to let the number

$$\rho(x, y) = \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot \frac{|x_k - y_k|}{1 + |x_k - y_k|}$$

be the distance between the two elements

$$x = (x_1, x_2, x_3, \ldots) \text{ and } y = (y_1, y_2, y_3, \ldots).$$

It is quite evident that the first two conditions which the distance must satisfy are satisfied. In order to establish the triangle inequality

$$\rho(x, z) \leqslant \rho(x, y) + \rho(y, z), \tag{4}$$

we note that the function

$$\varphi(t) = \frac{t}{1 + t}$$

increases $\left(\text{since } \varphi(t) = 1 - \frac{t}{1+t}\right)$ in the interval $0 \leqslant t < +\infty$. Hence if $z = (z_1, z_2, z_3, \ldots)$ is a third numerical sequence, then

$$\frac{|x_k - z_k|}{1 + |x_k - z_k|} \leqslant \frac{|x_k - y_k| + |y_k - z_k|}{1 + |x_k - y_k| + |y_k - z_k|}.$$

But

$$\frac{|x_k-y_k|}{1+|x_k-y_k|+|y_k-z_k|}\leqslant\frac{|x_k-y_k|}{1+|x_k-y_k|},$$

$$\frac{|y_k-z_k|}{1+|x_k-y_k|+|y_k-z_k|}\leqslant\frac{|y_k-z_k|}{1+|y_k-z_k|}.$$

Therefore

$$\frac{|x_k-z_k|}{1+|x_k-z_k|}\leqslant\frac{|x_k-y_k|}{1+|x_k-y_k|}+\frac{|y_k-z_k|}{1+|y_k-z_k|}.$$

Multiplying this inequality by 2^{-k} and summing with respect to k, we obtain (4). Thus, the linear space constructed turns out to be a metric space. But at the same time it is not normed.[2] This space is denoted by the letter s. It is not difficult to show that if

$$x^{(n)}=(\xi_1^{(n)},\xi_2^{(n)},\xi_3^{(n)},\ldots),\ x=(\xi_1,\xi_2,\xi_3,\ldots),$$

then a necessary and sufficient condition that the relation

$$\lim_{n\to\infty}x^{(n)}=x$$

hold is that for all k,

$$\lim_{n\to\infty}\xi_k^{(n)}=\xi_k.$$

Using this assertion, it is easy to verify that the space s is complete.

Space S. The space S is the space whose elements are all possible measurable functions defined on the same measurable set E of measure $mE>0$. In this connection, equivalent functions are not assumed to be distinct. Addition of the elements $f_1=f_1(x)$ and $f_2=f_2(x)$ of this space and the multiplication of the element $f=f(x)$ by the number a are defined, as in the case of the space C, by the formulas

$$f_1+f_2=f_1(x)+f_2(x),\ af=af(x).$$

Hence S is a linear space. It becomes a metric (but not a normed!) space if we set

$$\rho(f_1,f_2)=\int_E\frac{|f_1(x)-f_2(x)|}{1+|f_1(x)-f_2(x)|}\,dx.$$

Convergence, $\lim f_n=f$ [where $f_n=f_n(x)$, $f=f(x)$], *in this space is the convergence in measure of the sequence* $\{f_n(x)\}$ *to the function* $f(x)$. We leave the proof of this assertion to the reader; also left to the reader is the verification of the fact that the space S is complete.

§2. COMPACTNESS

Having established the concept of limit of a sequence of elements of a metric space we can take over, without any difficulty, the more important concepts of the theory of point sets into the theory of these spaces. So, for example, if A is a set of elements of a metric space and it is possible to select a sequence of distinct elements $\{x_n\}$ from A having as limit the element $x\in E$, we then say that x is a *limit element*, or a *limit point*,

[2] If we set $\|x\|=\rho(x,\theta)$, then the condition $\|ax\|=|a|\cdot\|x\|$ will not be fulfilled. For example, if $e_0=(1,1,1,\ldots)$, then $\|e_0\|=\frac{1}{2}$, $\|2e_0\|=\frac{2}{3}$.

of the set A. A set which contains all its limit points is said to be *closed.* The set $\bar{A}$ obtained from the set A by adjoining to it all its limit points is called the *closure* of the set A. The complement of a closed set is called an open set, and so on.

If x_0 is a fixed element of a metric space and r is a positive number, then the *open sphere* with center x_0 and radius r is the set $S_r(x_0)$ consisting of all points x of the space for which $\rho(x, x_0) < r$. Analogously, the set $\overline{S_r(x_0)}$ of those points x for which $\rho(x, x_0) \leqslant r$ is called the *closed sphere* with center x_0 and radius r. It is easy to show that the closed sphere is a closed set,[3] the open sphere is an open set, and the sphere $\overline{S_r(x_0)}$ is the closure of the sphere $S_r(x_0)$. With the aid of the concept of a sphere, we introduce in the usual manner the concept of *interior point* of a set and using the latter concept, it is easy to show that open sets are sets all of whose points are interior points. We shall not take time to discuss all these items in view of the fact that they do not contain any new ideas of principle.

A set A of elements in a metric space E is said to be *bounded* if it is completely contained in some sphere. In Chapter II we proved the Bolzano-Weierstrass theorem which states that every bounded infinite set of points on the real line has at least one limit point. At the very beginning of Chapter XI this result was carried over from the case of the real line to the case of an arbitrary Euclidean space. However, it does not carry over to *every* metric space. In fact, let us consider, say, the space L_2. For the set A, take some orthonormal system $\{\omega_k\}$ $[\omega_k = \omega_k(x)]$. Since for arbitrary k, $\|\omega_k\| = 1$, this set is bounded. But for $i \neq k$,

$$\|\omega_i - \omega_k\|^2 = \int_a^b [\omega_i^2(x) - 2\omega_i(x)\omega_k(x) + \omega_k^2(x)]\,dx = 2$$

and hence it is impossible to select any convergent sequence from A. Therefore, in the general theory of metric spaces, the concept of a bounded set no longer has the same fundamental meaning as in the case of Euclidean spaces. Its place is taken by the concept of a *compact* space.

DEFINITION 1. A non-void set A of elements in a metric space E is said to be *compact* if it is finite, or if every infinite subset A_0 of this set has at least one limit point.

Furthermore, we must note that the following theorem is valid.

THEOREM 1. *Every compact set is bounded.*

Assume that the set A is not bounded. Having fixed any point $x_0 \in E$, we can find a point $x_1 \in A$ such that $\rho(x_1, x_0) > 1$. Next we can find a point $x_2 \in A$ such that

$$\rho(x_2, x_0) > \rho(x_1, x_0) + 1.$$

Continuing this process, we obtain a sequence $x_1, x_2, x_3, \ldots$ of points of the set A such that

$$\rho(x_n, x_0) > \rho(x_1, x_0) + \rho(x_2, x_0) + \ldots + \rho(x_{n-1}, x_0) + 1.$$

Then, for $n > m$,

$$\rho(x_n, x_0) > \rho(x_m, x_0) + 1,$$

and since

---

[3]For this, it is sufficient to show that the metric function $\rho(x, y)$ is continuous, i.e. that the relations $x_n \to x$, $y_n \to y$ imply the relation $\rho(x_n, y_n) \to \rho(x, y)$. But this is clear. In fact, $\rho(x_n, y_n) \leqslant \rho(x_n, x) + \rho(x, y) + \rho(y, y_n)$ and for sufficiently large n, $\rho(x_n, y_n) < \rho(x, y) + \varepsilon$. On the other hand, $\rho(x, y) \leqslant \rho(x, x_n) + \rho(x_n, y_n) + \rho(y_n, y)$ and for sufficiently large n, $\rho(x, y) < \rho(x_n, y_n) + \varepsilon$.

$\rho(x_n, x_0) < \rho(x_n, x_m) + \rho(x_m, x_0)$, we have $\rho(x_n, x_m) > 1$

and therefore no convergent subsequence can be selected from the sequence $\{x_n\}$, i.e. $\{x_n\}$ has no limit points. But then the set A is not compact.

The theorem just proved can be considerably strengthened. To discuss this we need the following definition.

DEFINITION 2. Suppose A is a set of points in the metric space E, $B \subset A$ is a subset of A, and that $\varepsilon > 0$ is a positive number. We say that B is an ε-net in the set A if for every element $x \in A$ there exists an element $y \in B$ such that $\rho(x, y) < \varepsilon$.

THEOREM 2. *If A is a compact set, then there exists a finite ε-net in A for arbitrary $\varepsilon > 0$.*

It is not difficult to see that every set in which there is a finite ε-net is necessarily bounded.[4] Therefore, the present theorem is truly a strengthening of Theorem 1. Going over to the proof, let us assume the theorem is not true. Then for some $\varepsilon > 0$, no ε-net exists in the set A. Choose any element x_1 in A. Since it does not form an ε-net, we can find an element x_2 in A such that $\rho(x_1, x_2) \geqslant \varepsilon$. But the system of two elements x_1 and x_2 also does not form an ε-net. Therefore we can find an element x_3 in A such that

$$\rho(x_1, x_3) \geqslant \varepsilon, \quad \rho(x_2, x_3) \geqslant \varepsilon.$$

Continuing this line of reasoning, we convince ourselves that there exists a sequence $\{x_n\}$ of elements in A such that $\rho(x_n, x_m) \geqslant \varepsilon$ for $n \neq m$.

But then this sequence has no limit elements, which contradicts the fact that A is compact. This completes the proof of the theorem.

Thus, a necessary condition that the set A be compact is that there exist a finite ε-net in A for arbitrary $\varepsilon > 0$. It is natural to inquire whether this condition is also sufficient. It turns out that this is not so for an arbitrary metric space. For example, let the metric space be the set of rational numbers contained in the segment $[0, 1]$ with the usual definition of distance, $\rho(x, y) = |x - y|$.

We shall consider this space itself as some set contained in it. It is clear that this set contains a finite ε-net for arbitrary $\varepsilon > 0$ (the set of points $0, \frac{1}{n}, \frac{2}{n}, \ldots, 1$ where $n\varepsilon > 1$, for instance, could serve as such a net). Nonetheless, this set is not compact, because it is possible to select a sequence

$$x_n = \frac{1}{3}\left(1 + \frac{1}{n}\right)^n$$

from it having an irrational limit $\frac{e}{3}$, i.e. devoid of limit points in the space. The reader has very likely guessed that the unpleasant result obtained has at its source the incompleteness of the space considered. And, in fact, the following theorem is valid.

THEOREM 3. *Suppose E is a complete metric space and that A is an infinite subset of E. If there exists a finite ε-net in A for every $\varepsilon > 0$, then A is a compact set.*

Denote any infinite subset of the set A by P and let $\varepsilon > 0$. Consider the finite $\frac{\varepsilon}{2}$-net corresponding to the number $\frac{\varepsilon}{2}$. Let this $\frac{\varepsilon}{2}$-net consist of the points $y_1, y_2, \ldots, y_s$. Then

[4] In fact, let the ε-net consist of the elements $y_1, y_2, \ldots, y_s$. We select any element x_0 in the space and let $\max_{(i=1,2,\ldots,s)} \{\rho(y_i, x_0)\} = d$. It is clear that $A \subset S_r(x_0)$, where $r = d + \varepsilon$.

the set A, and hence also its subset P, are contained in the sum of a finite number of open spheres:

$$P \subset \sum_{i=1}^{s} S_{\frac{\varepsilon}{2}}(y_i).$$

But then at least one of these spheres, say $S_{\frac{\varepsilon}{2}}(y_m)$, contains an infinite number of points contained in the set P (because otherwise P would be finite). If x' and x'' are two points in the (infinite!) intersection $Q = S_{\frac{\varepsilon}{2}}(y_m) \cdot P$, then $\rho(x', x'') < \varepsilon$. In other words, we have proved the following fact: every infinite subset P of the set A contains an infinite subset Q of diameter[5] not greater than some prescribed number ε,

$$d(Q) < \varepsilon.$$

Noting this, we take any infinite subset A_0 of the set A. According to what has been proved, an infinite set A_1 of diameter $d(A_1) \leqslant 1$ can be selected from A_0. Further, an infinite set A_2 of diameter $d(A_2) \leqslant \frac{1}{2}$ is selected from A_1. Continuing this process we obtain a nested sequence of infinite sets

$$A_0 \supset A_1 \supset A_2 \supset \ldots,$$

such that $d(A_n) < \frac{1}{n}$. From each A_n, we select a point x_n, seeing to it that the point x_n is different from $x_1, x_2, \ldots, x_{n-1}$. If $m > n$, then $x_m \in A_m \subset A_n$ and, therefore, $\rho(x_n, x_m) < \frac{1}{n}$. From this it is clear that the sequence $\{x_n\}$ converges in itself. But then (by virtue of the fact the space E is complete) the limit $x = \lim x_n$ exists. This limit is a limit point of A_0. So, every infinite subset A_0 of the set A has limit points and this, of course, means that A is compact.

In a similar way, we can prove

THEOREM 4. *Let A be an infinite set of elements in the metric space E. A necessary, and in the case the space E is complete also sufficient, condition that the set A be compact is that for every $\varepsilon > 0$ there exist in E a compact set B_ε possessing the following property*[6]: *for arbitrary $x \in A$, a $y \in B_\varepsilon$ can be found such that $\rho(x, y) < \varepsilon$.*

The necessity of the condition is trivial because in case the set is compact, it itself can be taken for B_ε. Going over to the proof of the sufficiency of this condition, we shall show that if it is satisfied, from every infinite subset P of the set A, we can select an infinite set Q of diameter less than some arbitrary prescribed $\varepsilon > 0$. Having done this, we can conclude the proof exactly as in Theorem 3.

So, let P be an infinite subset of A and let ε be some positive number. We construct the set $B_{\frac{\varepsilon}{3}}$ corresponding to the number $\frac{\varepsilon}{3}$ and for every $x \in P$ we select an element $y(x)$ from $B_{\frac{\varepsilon}{3}}$ such that

$$\rho(x, y(x)) < \frac{\varepsilon}{3}.$$

Let D be the set of all $y(x)$, where $x \in P$.

---

[5] As always, the diameter $d(Q)$ of the set Q is the least upper bound of the numbers $\rho(x', x'')$, where x' and x'' are elements in Q.

[6] We cannot call B_ε an ε-net in A because we have not assumed that $B_\varepsilon \subset A$.

To begin with, assume D contains only a finite number of distinct elements. Then at least one of them, say y_0, can be represented as $y_0 = y(x)$ for an infinite set of elements $x \in P$. This set can be taken as the set Q sought because its diameter is less than $\frac{2}{3}\varepsilon$. But if the set D is infinite, then (since it is a subset of the *compact* set $B_{\frac{\varepsilon}{3}}$) a convergent sequence

$$y(x_1), y(x_2), y(x_3), \ldots$$

can be extracted from it, where the elements $x_1, x_2, x_3, \ldots$ are distinct.

Since the sequence $\{y(x_n)\}$ is convergent, we can find an N such that, for $n > N$, $m > N$, we have

$$\rho(y(x_n), y(x_m)) < \frac{\varepsilon}{3}.$$

But then the set $\{x_{N+1}, x_{N+2}, x_{N+3}, \ldots\}$ can be taken for the set Q sought. In fact, if $n > N, m > N$, then

$$\rho(x_n, x_m) \leqslant \rho(x_n, y(x_n)) + \rho(y(x_n), y(x_m)) + \rho(y(x_m), x_m) < \varepsilon$$

and

$$d(Q) \leqslant \varepsilon.$$

This completes the proof of the theorem.

Theorems 3 and 4 give *general criteria for compactness in complete spaces.* However, for many spaces which are frequently encountered, simpler tests for compactness have been found. We stop to consider some of them.

§3. CONDITIONS FOR COMPACTNESS IN CERTAIN SPACES

In this section, we shall establish conditions for compactness of a set lying in any of the four spaces s, l_p, C and L_p.

Space s. Elements of this space are numerical sequences $x = (\xi_1, \xi_2, \xi_3, \ldots)$. Since the number ξ_k, appearing the k-th place in this sequence, i.e. the k-th "coordinate" of the point x, varies from one sequence to another, it is a function of the sequence. Therefore, it is proper to denote this number by $\xi_k(x)$.

With the aid of this notation, we can formulate the following theorem.

THEOREM 1. *Let $A = \{x\}$, where $x = (\xi_1(x), \xi_2(x), \xi_3(x), \ldots)$, be an infinite set of points in the space s. A necessary and sufficient condition that A be compact is that there exist a sequence of numbers $M_1, M_2, M_3, \ldots$ such that for arbitrary $x \in A$, we have*

$$|\xi_k(x)| \leqslant M_k \qquad (k = 1, 2, 3, \ldots).$$

In fact, assume that such numbers M_k exist. Select from A a sequence of distinct points $x_1, x_2, x_3, \ldots$ and construct a numerical sequence from the *first coordinates* of the selected points:

$$\xi_1(x_1), \ \xi_1(x_2), \ \xi_1(x_3), \ldots . \tag{1}$$

Since $|\xi_1(x_n)| \leqslant M$, we have, by the Bolzano-Weierstrass theorem, that (1) contains a convergent subsequence,

$$\xi_1(x_1^{(1)}), \ \xi_1(x_2^{(1)}), \ \xi_1(x_3^{(1)}), \ldots; \ \lim_{n\to\infty} \xi_1(x_n^{(1)}) = \alpha_1.$$

Note that here $x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, \ldots$ is a sequence which is a subsequence of the sequence $x_1, x_2, x_3, \ldots$, where the relative position of its terms is exactly as in the sequence $x_1, x_2, x_3, \ldots$.

The sequence of *second* coordinates

$$\xi_2(x_1^{(1)}),\ \xi_2(x_2^{(1)}),\ \xi_2(x_3^{(1)}),\ldots$$

of the selected points $x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, \ldots$ is bounded (because $|\xi_2(x_n^{(1)})| \leqslant M_2$), and therefore it contains a convergent subsequence

$$\xi_2(x_1^{(2)}),\ \xi_2(x_2^{(2)}),\ \xi_2(x_3^{(2)}),\ldots;\ \lim_{n\to\infty} \xi_2(x_n^{(2)}) = \alpha_2.$$

In this connection, the sequence $x_1^{(2)}, x_2^{(2)}, x_3^{(2)}, \ldots$ is a subsequence of the sequence $x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, \ldots$ and the relative order of any two terms in either sequence is the same in both sequences.

Continuing this process, we arrive at the following infinite array of distinct points of the set A:

$$\begin{array}{l} x_1^{(1)},\ x_2^{(1)},\ x_3^{(1)},\ \ldots \\ x_1^{(2)},\ x_2^{(2)},\ x_3^{(2)},\ \ldots \\ x_1^{(3)},\ x_2^{(3)},\ x_3^{(3)},\ \ldots \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots \end{array}$$

where every row, except the first, is obtained from the preceding one by removing some of the terms without destroying the relative order of the remaining terms. The rows of the array are such that the finite limits

$$\lim_{n\to\infty} \xi_p(x_n^{(p)}) = \alpha_p \qquad (p = 1, 2, 3,\ \ldots)$$

exist.

Having obtained the indicated array, we consider the sequence of elements

$$y_1 = x_1^{(1)},\ y_2 = x_2^{(2)},\ y_3 = x_3^{(3)},\ \ldots\ ,$$

lying along its diagonal. This is a sequence of *distinct* points of the set A.

If we fix some positive integer p, then the elements $y_p, y_{p+1}, y_{p+2}, \ldots$ will be elements in the p-th row of our array:

$$y_n = x_{m(n,p)}^{(p)}, \qquad (n \geqslant p)$$

where, obviously, $m(n,p) \geqslant n$. But then

$$\xi_p(y_n) = \xi_p(x_{m(n,p)}^{(p)}) \qquad (n \geqslant p)$$

and hence

$$\lim_{n\to\infty} \xi_p(y_n) = \alpha_p.$$

The last relation shows that the subsequence $y_1, y_2, y_3, \ldots$ of the sequence $x_1, x_2, x_3, \ldots$ of points in A has the limit point

$$a = (\alpha_1,\ \alpha_2,\ \alpha_3, \ldots).$$

This means that every denumerable subset $\{x_n\}$ of the set A has a limit point and therefore A is compact. This proves the sufficiency of the condition of the theorem.

Proceeding now to establishing its necessity, we assume that this condition is not satisfied. Then for a least one k a number M_k cannot be found so that $|\xi_k(x)| \leqslant M_k$ for all $x \in A$. Fixing this k, we can guarantee that the set of numbers $\{\xi_k(x)\}$ is not bounded. Therefore, it is possible to select a sequence $x_1, x_2, x_3, \ldots$ from A such that we have

$$\lim_{n\to\infty} \xi_k(x_n) = \infty.$$

It is clear that the sequence $\{x_n\}$ does not have limit points.[7] This completes the proof of the theorem.

Space l_p $(p \geqslant 1)$. This space consists of all sequences $x = (\xi_1(x), \xi_2(x), \xi_3(x), \ldots)$ for which

$$\|x\| = \left[\sum_{k=1}^{\infty} |\xi_k(x)|^p\right]^{\frac{1}{p}} < +\infty.$$

The condition for compactness in this space is given by the following theorem:

THEOREM 2. *Let A be an infinite set of points in the space l_p. In order that A be compact it is necessary and sufficient that the following two conditions be satisfied:*

1) *there exist numbers $M_1, M_2, M_3, \ldots$ such that for all $x \in A$*

$$|\xi_k(x)| \leqslant M_k \qquad (k = 1, 2, 3, \ldots);$$

2) *the series*

$$\sum_{k=1}^{\infty} |\xi_k(x)|^p$$

converges uniformly on the set A.

We shall first prove the *sufficiency* portion of the theorem. Assuming that the conditions are satisfied, we first of all find an m such that for all $x \in A$,

$$\sum_{k=m+1}^{\infty} |\xi_k(x)|^p < 1.$$

Keeping this m fixed, we shall have for all $x \in A$ that

$$\sum_{k=1}^{m} |\xi_k(x)|^p \leqslant \sum_{k=1}^{m} M_k^p.$$

Hence, setting

$$\sum_{k=1}^{m} M_k^p + 1 = H,$$

we have

$$\sum_{k=1}^{\infty} |\xi_k(x)|^p < H$$

for all $x \in A$.

Having established this, we consider any infinite subset A_0 of the set A. Exactly as in Theorem 1, we shall show that a sequence $y_1, y_2, y_3, \ldots$ can be selected from A_0 such that the finite limits

$$\lim_{n\to\infty} \xi_k(y_n) = \alpha_k \qquad (k = 1, 2, 3, \ldots)$$

exist.

For arbitrary N,

$$\sum_{k=1}^{N} |\xi_k(y_n)|^p < H.$$

Taking the limit here for increasing n and fixed N, we find that

---

[7] If the sequence $\{x_n\}$ contained a subsequence $\{x_{n_i}\}$ having the limit a, then it would turn out that $\xi_k(x_{n_i}) \to \xi_k(a)$, whereas $\xi_k(x_{n_i}) \to \infty$.

$$\sum_{k=1}^{N} |\alpha_k|^p \leqslant H,$$

from which it follows, since N is arbitrary, that the sequence $a = (\alpha_1, \alpha_2, \alpha_3, \ldots)$ is included in the space l_p.

We shall show that

$$a = \lim_{n\to\infty} y_n. \tag{2}$$

To this end, we take $\varepsilon > 0$ and find an h such that

$$\sum_{k=h+1}^{\infty} |\xi_k(x)|^p < \left(\frac{\varepsilon}{3}\right)^p$$

for all $x \in A$.

Then, as above (i.e. considering first of all the sum $\sum_{k=h+1}^{N}$), we verify that

$$\sum_{k=h+1}^{\infty} |\alpha_k|^p \leqslant \left(\frac{\varepsilon}{3}\right)^p.$$

But

$$\left[\sum_{k=h+1}^{\infty} |\xi_k(y_n) - \alpha_k|^p\right]^{\frac{1}{p}} \leqslant \left[\sum_{k=h+1}^{\infty} |\xi_k(y_n)|^p\right]^{\frac{1}{p}} + \left[\sum_{k=h+1}^{\infty} |\alpha_k|^p\right]^{\frac{1}{p}}.$$

Hence

$$\left[\sum_{k=h+1}^{\infty} |\xi_k(y_n) - \alpha_k|^p\right]^{\frac{1}{p}} < \frac{2}{3}\varepsilon$$

and

$$\|y_n - a\| < \left[\sum_{k=1}^{h} |\xi_k(y_n) - \alpha_k|^p\right]^{\frac{1}{p}} + \frac{2}{3}\varepsilon.$$

Since for sufficiently large n,

$$\sum_{k=1}^{h} |\xi_k(y_n) - \alpha_k|^p < \left(\frac{\varepsilon}{3}\right)^p,$$

for these n we have

$$\|y_n - a\| < \varepsilon,$$

which proves (2) and hence the compactness of the set A.

The *necessity* of condition 1) is proved literally the same as the necessity of the condition in Theorem 1. We now assume that condition 2) is not satisfied. Then there exists an $\varepsilon > 0$ such that for every integer n, we can find an element x_n in A for which

$$\sum_{k=n+1}^{\infty} |\xi_k(x_n)|^p \geqslant \varepsilon. \tag{3}$$

It is not difficult to see that an infinite number of the elements x_n must be distinct. But at the same time $\{x_n\}$ contains no convergent subsequence.

In fact, if the sequence $x_{n_1}, x_{n_2}, x_{n_3}, \ldots$, where $n_1 < n_2 < n_3 < \ldots$, had a limit $a = (\alpha_1, \alpha_2, \alpha_3, \ldots)$, we could first find an h such that

$$\sum_{k=h+1}^{\infty} |\alpha_k|^p < \frac{\varepsilon}{2^p}.$$

Then for sufficiently large $i\,(i > i_0)$, we would have

$$\| x_n - a \| < \frac{\varepsilon}{2},$$

whence, a fortiori, for these i,

$$\sum_{k=n_i+1}^{\infty} | \xi_k(x_{n_i}) - \alpha_k |^p < \frac{\varepsilon}{2^p}.$$

This means, assuming $i > i_0$ so large that $n_i \geqslant h$, that we would have:

$$\left(\sum_{k=n_i+1}^{\infty} | \xi_k(x_{n_l}) |^p \right)^{\frac{1}{p}} \leqslant \left(\sum_{k=n_i+1}^{\infty} | \xi_k(x_{n_l}) - \alpha_k |^p \right)^{\frac{1}{p}} + \left(\sum_{k=n_i+1}^{\infty} | \alpha_k |^p \right)^{\frac{1}{p}} < \varepsilon^{\frac{1}{p}},$$

and this contradicts inequality (3). This completes the proof of the theorem.

As an example, we consider the set A consisting of "unit base vectors" of the space l_p, i.e. elements of the form

$$e_n = (0, \ldots, 0, 1, 0, 0, \ldots) \quad (1 \text{ in the } n\text{-th place}).$$

It is clear here that condition 1) of the theorem is satisfied. Further, the series

$$\sum_{k=1}^{\infty} | \xi_k(e_n) |^p$$

converges for every n because $\xi_k(e_n) = 0$ for $k \neq n$. But the convergence of this series is not uniform with respect to n, and the set A is not compact. The proof of this fact is quite straightforward. In fact,

$$\lim_{n \to \infty} \xi_k(e_n) = 0$$

and, therefore, only the point $\Theta = (0, 0, 0, \ldots)$ could be the unique limit point of A. But for arbitrary n,

$$\| e_n - \Theta \| = 1$$

and therefore there are no limit points in A.[8] It is useful to note, further, that $\| e_n \| = 1$ for all n, and therefore A is a bounded set. We have thus obtained still another example of a bounded set which is not compact.

Space C. We go over now to the consideration of the compactness condition in the space C, and give the following definition.

DEFINITION. Let a family $A = \{f(x)\}$ of continuous functions be defined on the segment $[a, b]$. If to every $\varepsilon > 0$ there corresponds a $\delta > 0$ such that for $| x'' - x' | < \delta$ (where x' and x'' lie in $[a, b]$), for an arbitrary function in A, we have

$$| f(x'') - f(x') | < \varepsilon,$$

we say that the functions of the family A are *equicontinuous.*

To elucidate what was stated above, we note that the presence of such a δ for every *individual* function $f(x)$ in A follows directly from the continuity of these functions. The essential point here is the possibility of choosing one δ for all functions of the family at once.

---

[8] It is still simpler to establish this if we note that for $m \neq n$, we have $\| e_n - e_m \| = \sqrt[p]{2}$.

THEOREM 3 (C. ARZELÀ-G. ASCOLI). *Let $A=\{f(x)\}$ be an infinite family of continuous functions defined on the segment $[a,b]$. If*

1) *all functions of the family are uniformly bounded, $|f(x)|\leqslant M$, and*

2) *these functions are equicontinuous,*

then A contains a uniformly convergent subsequence.

Proof. Denote the set of all rational numbers of the segment $[a,b]$ by E. This set is denumerable and therefore, on the basis of Lemma 1, §4, Chapt. VIII, A contains a sequence of *distinct* functions $f_1(x),f_2(x),f_3(x),\ldots$, which converges at all points of the set E. We shall show that this sequence converges not only on E but also on the entire segment $[a,b]$. In fact, let x be an irrational point of this segment. Taking $\varepsilon>0$, we find a rational point $r\in[a,b]$ such that for all functions of the family A, we have

$$|f(x)-f(r)|<\frac{\varepsilon}{3}.$$

The existence of such an r follows from the equicontinuity condition on the functions of the family. Since the numerical sequence $\{f_n(r)\}$ converges, to the ε chosen there corresponds an N such that

$$|f_n(r)-f_m(r)|<\frac{\varepsilon}{3}$$

provided $n>N, m>N$.

But then for these n and m, we have

$$|f_n(x)-f_m(x)|\leqslant|f_n(x)-f_n(r)|+|f_n(r)-f_m(r)|+|f_m(r)-f_m(x)|<\varepsilon.$$

Thus, the sequence $\{f_n(x)\}$ converges in itself and hence it has a finite limit.

We have thus proved the existence of the limit

$$\lim_{n\to\infty} f_n(x)=\varphi(x)$$

for all $x\in[a,b]$. It is easy to see that the limit function $\varphi(x)$ is continuous. In fact, taking $\varepsilon>0$ and finding a $\delta>0$ such that the inequality $|x''-x'|<\delta$ implies the inequality $|f(x'')-f(x')|<\varepsilon$ for all functions $f(x)$ of the family, we fix points x' and x'' such that $|x''-x'|<\delta$ and pass to the limit in the inequality

$$|f_n(x'')-f_n(x')|<\varepsilon,$$

which brings us to an analogous inequality

$$|\varphi(x'')-\varphi(x')|\leqslant\varepsilon$$

for the limit function.

It remains to verify that the convergence of $f_n(x)$ to $\varphi(x)$ is *uniform* with respect to x.

To do this, note first of all that the difference $f_n(x)-\varphi(x)$, as well as the functions $f_n(x)$, are equicontinuous. Therefore, taking $\varepsilon>0$, we can find a $\delta>0$ such that for $|x''-x'|<\delta$ and for arbitrary n, we have

$$|\{f_n(x'')-\varphi(x'')\}-\{f_n(x')-\varphi(x')\}|<\frac{\varepsilon}{2}. \tag{4}$$

Having found such a δ, we take a natural number s so large that

$$\frac{b-a}{s}<\delta,$$

and put

$$z_0 = a, z_1 = a + \frac{b-a}{s}, \; z_2 = a + 2\frac{b-a}{s}, \ldots, \; z_s = b.$$

At every point z_i, we have

$$\lim_{n \to \infty} f_n(z_i) = \varphi(z_i)$$

and, therefore, we can find N_i such that for $n > N_i$ we have

$$|f_n(z_i) - \varphi(z_i)| < \frac{\varepsilon}{2}. \tag{5}$$

Let $N = \max\{N_0, N_1, \ldots, N_s\}$. If $n > N$, then for arbitrary $x \in [a, b]$, we have

$$|f_n(x) - \varphi(x)| < \varepsilon. \tag{6}$$

In fact, taking any $x \in [a, b]$, we can find an i such that $|z_i - x| < \frac{b-a}{s} < \delta$, and then (6) will follow from (4) and (5). Since the number N does not depend on the choice of x, the theorem is completely proved.

THEOREM 4. *A necessary and sufficient condition that an infinite set A of elements of the space C be compact is that the functions in A be uniformly bounded and equicontinuous.*

The *sufficiency* of the condition of the theorem is proved in the Arzelà-Ascoli theorem. The *necessity* of the uniform boundedness condition on the functions $f(x)$ in A follows from Theorem 1 of the preceding section, because the boundedness of the set A, which, in accordance with this theorem, follows from its compactness, also implies that the functions in A are uniformly bounded. It remains to establish the necessity of the equicontinuity condition. Assume that this condition is not satisfied. Then for some $\varepsilon_0 > 0$, we can find a sequence of functions $\{f_n(x)\}$ in A and two sequences $\{x_n\}$ and $\{y_n\}$ of points in $[a, b]$ possessing the following properties:

$$\lim (y_n - x_n) = 0, |f_n(x_n) - f_n(y_n)| \geqslant \varepsilon_0. \tag{7}$$

The sequence $\{f_n(x)\}$ must contain an infinite number of *distinct* elements of A. But at the same time, A contains no uniformly convergent subsequence. In fact, suppose such subsequence has been selected from $\{f_n(x)\}$. Without loss of generality, we may assume that this is the sequence $\{f_n(x)\}$ itself, because this is what everything reduces to by a simple change of notation. So, besides relation (7), it is known that the sequence $\{f_n(x)\}$ converges uniformly to some (obviously continuous) function $\varphi(x)$.

For the $\varepsilon_0 > 0$, mentioned above, there exists a $\delta > 0$ such that $|\varphi(x'') - \varphi(x')| < \frac{\varepsilon_0}{3}$ for $|x'' - x'| < \delta$. On the other hand, it is possible to find an N such that for $n > N$ and arbitrary $x \in [a, b]$, we have

$$|f_n(x) - \varphi(x)| < \frac{\varepsilon_0}{3}.$$

Take $n > N$, and subject it further to the condition that $|y_n - x_n| < \delta$. For this n, it turns out that

$$|f_n(x_n) - f_n(y_n)| \leqslant |f_n(x_n) - \varphi(x_n)| + |\varphi(x_n) - \varphi(y_n)| + |\varphi(y_n) - f_n(y_n)| < \varepsilon_0.$$

But this contradicts condition (7). Thus, if the equicontinuity condition on the functions in the set A is not satisfied, then A is not compact. This completes the proof of the theorem.

We introduce an example[9] of how Theorem 3 is applied in classical analysis. Consider the differential equation

$$\frac{dy}{dx}=f(x,y), \tag{8}$$

the right member of which is defined and continuous on some closed rectangle R.

THEOREM 5. *If (x_0, y_0) is an interior point of R, then at least one integral curve of equation (8) passes through this point.*

Proof. Since $f(x,y)$ is continuous in R, it is bounded there. Let

$$|f(x,y)|< M.$$

Draw the straight lines I and II through (x_0, y_0) with slopes M and $-M$, respectively. Construct, further, two straight lines $x=a$ and $x=b, a<x_0<b$, choosing the numbers a and b so close to x_0 that the triangles ABC and ADE, bounded by the lines $x=a$ and $x=b$ and the lines I and II drawn earlier, lie entirely in the rectangle R (see Fig. 9).

Having made this construction, we subdivide the segment $[a,b]$ by the points $z_0=a< z_1< z_2< \ldots < z_s=b$ into s parts, taking care that the point x_0 coincide with one of the division points, by letting, for instance, $x_0=z_m$. Through (x_0,y_0), draw a straight line with slope $f(x_0,y_0)$. This line will lie within the vertical angles BAC and DAE. Moving from the point (x_0,y_0) to the right along the drawn line, we encounter

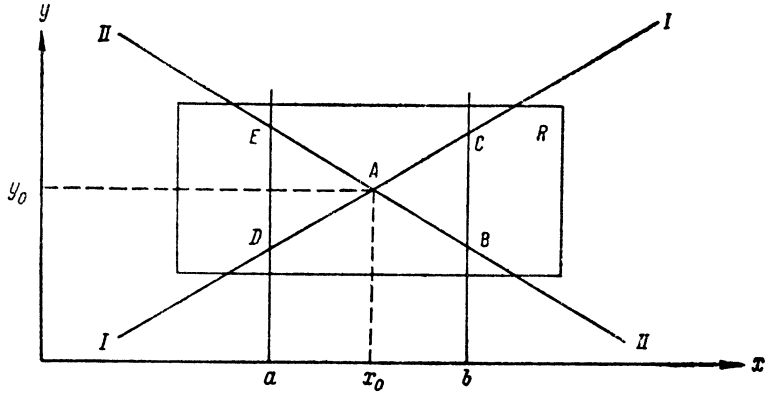


Fig. 9

the point (z_{m+1}, u_{m+1}) on it. Assuming $m+1<s$, we draw a straight line through this point with slope $f(z_{m+1}, u_{m+1})$ and consider the segment on it which lies over the segment $[z_{m+1}, z_{m+2}]$. This segment lies entirely in the triangle ABC. If $m+2< s$, then through the right end (z_{m+2}, u_{m+2}) of the segment obtained, we draw a new line with slope $f(z_{m+2}, u_{m+2})$, and we continue this process until we obtain a right endpoint of one of the constructed segments which lies on the line $x=b$. We carry out an analogous construction moving to the left from the point (x_0,y_0).

As a result, we obtain a polygonal arc having vertices in the points

$$(z_0,u_0),(z_1,u_1),\ldots,(z_s,u_s),$$

where $z_m=x_0, u_m=y_0$ and such that the slope of the link joining the vertices (z_k,u_k)

[9]The discussion here follows I. G. Petrovsky (see his *Lectures on the Theory of Ordinary Differential Equations*, GONTI, 1939, §12).

and (z_{k+1}, u_{k+1}) is $f(z_k, u_k)$ for $k \geqslant m$ and $f(z_{k+1}, u_{k+1})$ for $k < m$. This polygonal arc is called the *Euler polygonal arc*. It is important to note that it does not extend outside of the triangles ABC and ADE (this is proved most simply by induction in the very construction of the polygonal arc, going successively from one link of the polygonal arc to the next). Consider the function $\varphi(x)$ whose graph is the Euler polygonal arc. This function is bounded because its graph does not extend beyond the bounds of the rectangle R. It is also not difficult to see that $\varphi(x)$ satisfies the Lipschitz condition

$$|\varphi(x'') - \varphi(x')| \leqslant M|x'' - x'|.$$

We call $\varphi(x)$ an *Euler function* and say that it corresponds to the subdivision of $[a, b]$ carried out above. Imagine all possible Euler functions which can be obtained by varying the method of subdividing $[a, b]$. In accordance with what has been stated above, these functions are uniformly bounded and equicontinuous. This means the set of these functions is compact in C.

We now consider a sequence of subdivisions of $[a, b]$ by the points $z_0, z_1, \ldots, z_s$ such that

$$\max [z_{k+1} - z_k] \to 0.$$

To these methods of subdivision, there correspond certain Euler functions. Select from this sequence of Euler functions a uniformly convergent subsequence $\varphi_1(x), \varphi_2(x), \varphi_3(x), \ldots$. Let

$$\lim_{n\to\infty} \varphi_n(x) = \psi(x).$$

Obviously, $\psi(x_0) = y_0$, so that the curve $y = \psi(x)$ passes through the point (x_0, y_0). We shall prove this is an integral curve of equation (8).

To this end, we denote by $z_0^{(n)}, \ldots, z_{s_n}^{(n)}$ the division points of that method of subdivision to which the function $\varphi_n(x)$ corresponds. Recall that

$$\lim_{n\to\infty} [\max_k \{z_{k+1}^{(n)} - z_k^{(n)}\}] = 0. \tag{9}$$

Take any point x lying in the half-open interval $(x_0, b]$. Let

$$z_p^{(n)} < x \leqslant z_{p+1}^{(n)} \qquad [p = p(n)].$$

Then, retaining the former notation that z_m $[m = m(n)]$ is the division point coinciding with x_0, we have

$$\begin{aligned}
&\varphi_n(z_{m+1}^{(n)}) = \varphi_n(z_m^{(n)}) + f(x_0, \varphi_n(x_0))(z_{m+1}^{(n)} - x_0)\\
&\varphi_n(z_{m+2}^{(n)}) = \varphi_n(z_{m+1}^{(n)}) + f(z_{m+1}^{(n)}, \varphi_n(z_{m+1}^{(n)}))(z_{m+2}^{(n)} - z_{m+1}^{(n)})\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&\varphi_n(z_p^{(n)}) = \varphi_n(z_{p-1}^{(n)}) + f(z_{p-1}^{(n)}, \varphi_n(z_{p-1}^{(n)}))(z_p^{(n)} - z_{p-1}^{(n)})\\
&\varphi_n(x) = \varphi_n(z_p^{(n)}) + f(z_p^{(n)}, \varphi_n(z_p^{(n)}))(x - z_p^{(n)}).
\end{aligned}$$

Adding these equalities, we find that

$$\varphi_n(x) = y_0 + \sum_{k=m}^{p-1} f(z_k^{(n)}, \varphi_n(z_k^{(n)}))(z_{k+1}^{(n)} - z_k^{(n)}) + f(z_p^{(n)}, \varphi_n(z_p^{(n)}))(x - z_p^{(n)}).$$

The function $f(x, y)$ is uniformly continuous. This means that if we take an arbitrary $\varepsilon > 0$, we can find a $\delta > 0$ such that for $|y'' - y'| < \delta$,

$$|f(x,y'')-f(x,y')|<\varepsilon$$

for arbitrary $x\in[a,b]$. Without loss of generality, we may assume that $\delta\leqslant\varepsilon$. We shall assume that n is so large that for all $x\in[a,b]$ we have

$$|\varphi_n(x)-\psi(x)|<\delta.$$

Then the sum

$$\sum_{k=m}^{p-1} f(z_k^{(m)},\varphi_n(z_k^{(n)}))(z_{k+1}^{(n)}-z_k^{(n)})+f(z_p^{(n)},\varphi_n(z_p^{(n)}))(x-z_p^{(n)})$$

will differ from the analogous sum

$$\sigma_n=\sum_{k=m}^{p-1} f(z_k^{(n)},\psi(z_k^{(n)}))(z_{k+1}^{(n)}-z_k^{(n)})+f(z_p^{(n)},\psi(z_p^{(n)}))(x-z_p^{(n)})$$

by less than $\varepsilon(x-x_0)$. From this it follows that for the sufficiently large n indicated above, we have

$$|\psi(x)-y_0-\sigma_n|<\varepsilon(1+x-x_0). \tag{10}$$

Noting that the sum σ_n is a Riemann sum for the integral

$$\int_{x_0}^{x} f[z,\psi(z)]\,dx,$$

taking (9) into account, and taking the limit in (10) for increasing n, we find that

$$|\psi(x)-y_0-\int_{x_0}^{x} f[z,\psi(z)]\,dz|\leqslant\varepsilon(1+x-x_0).$$

Hence, since ε is arbitrary, it follows that

$$\psi(x)=y_0+\int_{x_0}^{x} f[z,\psi(z)]\,dz. \tag{11}$$

Differentiating this equality with respect to the argument x, we obtain[10]

$$\psi'(x)=f[x,\psi(x)] \tag{12}$$

for arbitrary $x\in[x_0,b]$.

We establish (12) for $a\leqslant x\leqslant x_0$ analogously. This completes the proof of the theorem.

Remark 1. The reader's attention is called to the fact that if the rectangle R is defined by the inequalities $R(x_0-A\leqslant x\leqslant x_0+A, y_0-B\leqslant y\leqslant y_0+B)$, then the proof just carried out guarantees the existence of a solution on the segment $[x_0-\delta, x_0+\delta]$, where

$$\delta=\min\left\{A,\frac{B}{M}\right\}.$$

Remark 2. Under the conditions of Theorem 5, the integral curve passing through (x_0,y_0) is not necessarily *unique*. For example, two integral curves $y=0$ and $y=x^3$ of the equation $\frac{dy}{dx}=3\sqrt[3]{y^2}$ pass through $(0,0)$.

---

[10]Although (11) was proved only for $x>x_0$, (12) follows from (11) also for $x=x_0$. It is evident that for $x=x_0$ and $x=b$, one must consider only one-sided derivatives.

Space L_p. In conclusion of §3, we consider the conditions for compactness in the space L_p. These conditions, for $p > 1$, were found by Kolmogorov in 1931. Two years later, A. N. Tulaikov[11] proved that the Kolmogorov conditions also turn out to be necessary and sufficient for the case $p = 1$.

To establish the results which interest us, some auxiliary considerations will be needed.

LEMMA 1. *Let a summable function $\varphi(t)$ be defined on $[a, b]$. Extend its definition to the segment $[a-h, b+h]$, by setting $\varphi(t) = 0$ for $t \bar{\in} [a, b]$, and construct a new function $\varphi_h(x)$, defining it on the segment $[a, b]$ by the formula*

$$\varphi_h(x) = \frac{1}{2h} \int_{x-h}^{x+h} \varphi(t)\, dt.$$

The so-called Steklov function $\varphi_h(x)$ is continuous on $[a, b]$.

In fact,

$$\varphi_h(x) = \frac{1}{2h}[F(x+h) - F(x-h)],$$

where

$$F(x) = \int_{a-h}^{x} \varphi(t)\, dt,$$

and $F(x)$, as is known, is a continuous function.

LEMMA 2. *Using the same notation as in Lemma 1, above, we have*

$$\int_a^b |\varphi_h(x)|\, dx \leqslant \int_a^b |\varphi(x)|\, dx. \tag{13}$$

Proof. Assume first that $\varphi(x)$ is non-negative. Consider the function $\varphi(z+t)$ in the rectangle $R(a \leqslant t \leqslant b, -h \leqslant z \leqslant +h)$. According to Theorem 2, §4, Chapt. XII, we have [12]

$$\int_a^b dt \int_{-h}^{+h} \varphi(z+t)\, dz = \int_{-h}^{+h} dz \int_a^b \varphi(z+t)\, dt. \tag{14}$$

Since

$$\int_{-h}^{+h} \varphi(z+t)\, dz = \int_{t-h}^{t+h} \varphi(x)\, dx = 2h\varphi_h(t),$$

the integral in the left member of (14) reduces to

---

[11] *Zur Kompaktheit im Raum L_p für p = 1*, Göttinger Nachrichten, 1933, 167-170.

[12] The measurability of $\varphi(z+t)$ as a function of one of the arguments t and z, for fixed second argument, is clear. We shall establish that it is also measurable in R if we consider it as a function of *two* arguments. For arbitrary c, the set of those $x \in [a-h, b+h]$, for which $\varphi(x) > c$, is measurable. But then (see Lemma 1, §2, Chapt. XII) the set A of points (x, y) in the rectangle $(a - h \leqslant x \leqslant b + h, -h \leqslant y \leqslant h)$, for which $\varphi(x) > c$, is measurable. Suppose E is the set of points (t, z) into which A is mapped under the affine transformation of the plane according to the formulas $t = x - y, z = y$. It is measurable and now it remains to take into consideration the fact that

$$R[\varphi(z+t) > c] = RE.$$

$$2h\int_a^b \varphi_h(x)\,dx.$$

But the integral in the right member of (14) transforms into the form

$$\int_{-h}^{+h} dz \int_{a+z}^{b+z} \varphi(x)\,dx,$$

and it remains to note[13] that

$$\int_{a+z}^{b+z} \varphi(x)\,dx \leqslant \int_a^b \varphi(x)\,dx \tag{15}$$

in order to arrive at (13).

We shall now omit the assumption that $\varphi(x) \geqslant 0$ and denote the Steklov function for $|\varphi(x)|$ by $\bar{\varphi}_h(x)$. Then

$$|\varphi_h(x)| = \left|\frac{1}{2h}\int_{x-h}^{x+h} \varphi(t)\,dt\right| \leqslant \frac{1}{2h}\int_{x-}^{x+h} |\varphi(t)|\,dt = \bar{\varphi}_h(x)$$

and

$$\int_a^b |\varphi_h(x)|\,dx \leqslant \int_a^b \bar{\varphi}_h(x)\,dx. \tag{16}$$

But, according to what we have already proved,

$$\int_a^b \bar{\varphi}_h(x)\,dx \leqslant \int_a^b |\varphi(x)|\,dx,$$

from which, in combination with (16), (13) follows.

LEMMA 3. *If $\varphi(t) \in L_p$ for $p \geqslant 1$, then*

$$\|\varphi_h\| \leqslant \|\varphi\|. \tag{17}$$

Here, as usual,

$$\|\varphi\| = \sqrt[p]{\int_a^b |\varphi(x)|^p\,dx}.$$

It is useful to note that, in virtue of the continuity of the Steklov function, $\varphi_h(x)$, it belongs to L_p.

Lemma 3 reduces to Lemma 2 in the case when $p = 1$. Therefore, we may assume that $p > 1$. But then, in virtue of the Hölder inequality, we have

$$\left|\int_{x-h}^{x+h} \varphi(t)\,dt\right| \leqslant \left(\int_{x-h}^{x+h} |\varphi(t)|^p\,dt\right)^{\frac{1}{p}} \left(\int_{x-h}^{x+h} dt\right)^{\frac{1}{q}}, \quad \left(\frac{1}{q}+\frac{1}{p}=1\right)$$

---

[13] For $z = 0$, the integrals in (15) are equal. But if, for example, $z > 0$, then

$$\int_{a+z}^{b+z} \varphi(t)\,dt = \int_{a+z}^{b} \varphi(t)\,dt \leqslant \int_a^b \varphi(t)\,dt,$$

inasmuch as $\varphi(t) \geqslant 0$.

or, equivalently,

$$|\varphi_h(x)|^p \leqslant \frac{1}{2h}\int_{x-h}^{x+h} |\varphi(t)|^p dt. \tag{18}$$

The right member of this inequality is nothing else than the Steklov function for the function $|\varphi(x)|^p$. Denoting it by $\bar{\varphi}_h(x)$, we can rewrite inequality (18) in the form

$$|\varphi_h(x)|^p \leqslant \bar{\varphi}_h(x).$$

It follows that

$$\int_a^b |\varphi_h(x)|^p dx \leqslant \int_a^b \bar{\varphi}_h(x)\, dx,$$

and in virtue of Lemma 2, we shall have

$$\int_a^b \bar{\varphi}_h(x)\, dx \leqslant \int_a^b |\varphi(x)|^p dx.$$

Now (17) follows from these inequalities.

LEMMA 4. *If* $p \geqslant 1$ *and* $\varphi(x) \in L_p$, *then*

$$\lim_{h \to 0} \int_a^b |\varphi_h(x) - \varphi(x)|^p dx = 0. \tag{19}$$

In fact, suppose first of all that the function $\varphi(x)$ is continuous. If $a < x < b$ and h is so small that $[x-h, x+h] \subset [a, b]$, then, by the mean value theorem,

$$\varphi_h(x) = \frac{1}{2h}\int_{x-h}^{x+h} \varphi(t)\, dt = \varphi(\xi),$$

where $\xi \in [x-h, x+h]$. It is clear from this that if $x \in (a, b)$, then

$$\lim_{h \to 0} \varphi_h(x) = \varphi(x)$$

and hence the function under the integral sign in (19) tends to zero almost everywhere in $[a, b]$ [more precisely, everywhere in the open interval (a, b)]. On the other hand, $\varphi(x)$, being continuous, is bounded. But if $|\varphi(x)| \leqslant M$, then $|\varphi_h(x)|$ is also $\leqslant M$. This means that the function under the integral sign in (19) is bounded by a number which is independent of the parameter. It follows from this that it is possible to pass to the limit under the integral sign, which completes the proof of formula (19).

Suppose now that $\varphi(x)$ is an arbitrary function in L_p. Choosing $\varepsilon > 0$, we find a continuous function $\psi(x)$ such that

$$\|\psi - \varphi\| < \varepsilon. \tag{20}$$

We denote the Steklov function for the function $\psi(x)$ by $\psi_h(x)$ and note that the Steklov function for the difference $\varphi(x) - \psi(x)$ is the difference $\varphi_h(x) - \psi_h(x)$; then, according to Lemma 3, it follows from (20) that

$$\|\psi_h - \varphi_h\| < \varepsilon.$$

But

$$\|\varphi_h - \varphi\| \leqslant \|\varphi_h - \psi_h\| + \|\psi_h - \psi\| + \|\psi - \varphi\|.$$

Hence

$$\|\psi_h - \varphi\| < 2\varepsilon + \|\psi_h - \psi\|,$$

and by what we have already proved, we shall have $\|\psi_h - \psi\| < \varepsilon$ for sufficiently small h, so that $\|\varphi_h - \varphi\| < 3\varepsilon$ for these h. This completes the proof of the lemma.

THEOREM 6 (A. N. KOLMOGOROV). *Suppose $A = \{f(x)\}$ is a set of functions in $L_p (p > 1)$. In order that this set be compact, it is necessary and sufficient that both of the following conditions be satisfied:*

1) *the set A is bounded in L_p;*

2) *the difference $\|f_h - f\|$ tends to zero uniformly, as $h \to 0$, relative to the choice of functions $f(x)$ in A.*

The *necessity* of condition 1) is clear because every compact set is bounded. In order to prove the necessity of condition 2), we shall assume that it is not satisfied. Then there exists an $\varepsilon_0 > 0$ and two sequences $\{h_n\}$ and $\{f^{(n)}(x)\}$, where the $h_n > 0$ are numbers, and the $f^{(n)}(x)$ are functions in A, such that

$$\lim h_n = 0, \qquad \|f^{(n)}_{h_n} - f^{(n)}\| \geqslant \varepsilon_0.$$

We shall show that $\{f^{(n)}(x)\}$ contains no convergent subsequence. Let us assume, on the contrary, that such a subsequence can be selected. Without loss of generality, we may assume that this is the sequence $\{f^{(n)}(x)\}$ itself, because a simple change in notation reduces the situation to this. Thus, suppose the sequence $\{f^{(n)}(x)\}$ tends to the function $g(x)$ (relative to the metric in L_p). We then have

$$\varepsilon_0 \leqslant \|f^{(n)}_{h_n} - f^{(n)}\| \leqslant \|f^{(n)}_{h_n} - g_{h_n}\| + \|g_{h_n} - g\| + \|g - f^{(n)}\|,$$

where the notation used is self-evident. Hence, by Lemma 3,

$$\varepsilon_0 \leqslant 2\|g - f^{(n)}\| + \|g_{h_n} - g\|.$$

But (by Lemma 4), for sufficiently large n, we have $\|g_{h_n} - g\| < \frac{\varepsilon_0}{2}$ and, hence, for such n

$$\|g - f^{(n)}\| > \frac{\varepsilon_0}{4},$$

which contradicts our assumption that $f^{(n)}(x)$ tends to $g(x)$. The necessity of the conditions in the theorem is proved.[14]

We shall now assume that conditions 1) and 2) are satisfied. Then for arbitrary $f(x) \in A$, we have

$$\|f\| = \left(\int_a^b |f(x)|^p dx \right)^{\frac{1}{p}} < M,$$

where the constant M does not depend on the choice of $f(x)$. Fix some $h > 0$ and consider the set of Steklov functions $A_h = \{f_h(x)\}$ constructed, with this h, for all functions $f(x) \in A$. By Hölder's inequality, we have for arbitrary $f_h(x)$ that

$$|f_h(x)| \leqslant \frac{1}{2h} \left(\int_{x-h}^{x+h} |f(t)|^p dt \right)^{\frac{1}{p}} \left(\int_{x-h}^{x+h} dt \right)^{\frac{1}{q}} < \frac{M}{(2h)^{\frac{1}{p}}} \left(\frac{1}{p} + \frac{1}{q} = 1 \right).$$

[14] In this connection, the line of reasoning followed did not depend on the fact that $p > 1$; it is also valid for $p = 1$.

This means that all the functions in A_h are uniformly bounded. We shall show that they are equicontinuous. In fact, for any one of them, we have

$$f_h(x')-f_h(x)=\frac{1}{2h}\left\{\int_{x+h}^{x'+h} f(t)\,dt-\int_{x-h}^{x'-h} f(t)\,dt\right\} \tag{21}$$

where $a \leqslant x < x' \leqslant b$. But

$$\left|\int_{x+h}^{x'+h} f(t)\,dt\right| \leqslant \left(\int_{x+h}^{x'+h} |f(t)|\,dt\right)^{\frac{1}{p}} \left(\int_{x+h}^{x'+h} dt\right)^{\frac{1}{q}} < M(x'-x)^{\frac{1}{q}}.$$

A similar estimate is also valid for the second integral in the curly brackets in (21). Hence

$$|f_h(x')-f_h(x)| < \frac{M}{h}(x'-x)^{\frac{1}{q}},$$

from which the equicontinuity of the functions $\{f_h(x)\}$ is clear.

Thus, for arbitrary $h>0$, the set A_h is compact in C. Therefore, it is a fortiori compact in L_p. But then the compactness of the set A follows directly from condition 2) and Theorem 4, §2.

Remark. V. N. Sudakov[15] proved that the second condition in Theorem 6 implies the first.

§4. BANACH'S "FIXED-POINT PRINCIPLE" AND SOME OF ITS APPLICATIONS

Consider a system of n linear equations in n unknowns. It can always be written in the form

$$\left.\begin{aligned} x_1 &= a_{1,1}x_1 + a_{1,2}x_2 + \ldots + a_{1,n}x_n + b_1 \\ x_2 &= a_{2,1}x_1 + a_{2,2}x_2 + \ldots + a_{2,n}x_n + b_2 \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n &= a_{n,1}x_1 + a_{n,2}x_2 + \ldots + a_{n,n}x_n + b_n \end{aligned}\right| \tag{1}$$

We shall show that it is possible to look at the problem of solving system (1) from an entirely new point of view. For this purpose, we consider the system of n equations

$$\left.\begin{aligned} y_1 &= a_{1,1}x_1 + a_{1,2}x_2 + \ldots + a_{1,n}x_n + b_1 \\ y_2 &= a_{2,1}x_1 + a_{2,2}x_2 + \ldots + a_{2,n}x_n + b_2 \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n &= a_{n,1}x_1 + a_{n,2}x_2 + \ldots + a_{n,n}x_n + b_n \end{aligned}\right| \tag{2}$$

For any system of numbers $(x_1, x_2, \ldots, x_n)$ which we substitute into the right members of equations (2), we obtain certain numbers $(y_1, y_2, \ldots, y_n)$ in the left members. Thus, equations (2) define an operation, or, as we more frequently say, an operator, which transforms the number complex $(x_1, x_2, \ldots, x_n)$ into the complex $(y_1, y_2, \ldots, y_n)$. In other words, we can say that equations (2) define an operator which maps points of n-dimensional Euclidean space R_n into points of the same space. With this approach, the problem of solving system (1) is that of finding a point of the space R_n which is mapped into itself by the operator (2), i.e. a "fixed" point of the transformation.

[15] *Concerning Criteria for Compactness in Function Spaces*, Uspekhi Mat. Nauk, 12, No. 3, 1957.

It turns out that this treatment of the problem is possible not only with regard to linear algebraic equations, but also with respect to differential equations, integral equations, systems of such equations, and many other problems of analysis. A very fruitful method[16] of investigating all such problems was initiated in this connection; this is the so-called "fixed point method." Its theory consists in the investigation of different types of operators and establishing the conditions under which these operators leave definite points fixed. We shall restrict ourselves to the consideration of only very special conditions of this type, referring the reader to the article, mentioned above, by V. V. Nemytsky, for more detailed information.

DEFINITION 1. Let E and E_1 be metric spaces and let U be a rule which associates some point $y \in E_1$ with every point $x \in E$. Such a rule is called an *operator* which is defined in the space E and maps E into the space E_1. If $y \in E_1$ is a point which the operator U assigns to the point $x \in E$, we write $y = U(x)$ and we call y a *value* of the operator.

DEFINITION 2. Let the operator U map the metric space E into itself. If there exists a number $q, 0 \leqslant q \leqslant 1$, such that for arbitrary points x and x' of the space E, we have

$$\rho(U(x), U(x')) \leqslant q \cdot \rho(x, x'), \tag{3}$$

then we call U a *rapprochement operator* or a *contraction operator*.

THEOREM (S. S. BANACH). *If a contraction operator U is defined on a complete metric space E, then there exists a unique point x^* in this space for which*

$$U(x^*) = x^*. \tag{4}$$

This point is called the *fixed point* of the operator U. To prove this theorem, we take an arbitrary point x_0 in the space E and construct the sequence

$$x_1 = U(x_0), \quad x_2 = U(x_1), \quad x_3 = U(x_2), \ldots .$$

We shall show that this sequence converges in itself. In fact, for $n \geqslant 1$, we have

$$\rho(x_{n+1}, x_n) = \rho(U(x_n), \ U(x_{n-1})) \leqslant q \cdot \rho(x_n, x_{n-1}).$$

From this it follows that

$$\rho(x_{n+1}, x_n) \leqslant aq^n,$$

where for brevity we have set $a = \rho(x_1, x_0)$. Having noted this, we choose any natural number N and let $n > N, m > N$. For definiteness, we shall assume that $m > n$. Then

$$\rho(x_n, x_m) \leqslant \rho(x_n, x_{n+1}) + \rho(x_{n+1}, x_{n+2}) + \ldots + \rho(x_{m-1}, x_m).$$

Hence

$$\rho(x_n, x_m) \leqslant aq^n + aq^{n+1} + \ldots + aq^{m-1} \leqslant \frac{aq^n}{1-q}$$

and, a fortiori,

$$\rho(x_n, x_m) \leqslant \frac{aq^N}{1-q}.$$

Since q^N tends to zero with increasing N, the convergence of the sequence $\{x_n\}$ into itself is proved. But the space E is complete. This means that the limit

---

[16] A very thorough discussion of this method is given in the article *The Fixed Point Method in Analysis* by V. V. Nemytsky, Uspekhi Mat. Nauk, No. 1, 1936.

$$\lim_{n\to\infty} x_n = x^*$$

exists.

By (3), we have

$$\rho(x_{n+1}, U(x^*)) = \rho(U(x_n), U(x^*)) \leqslant q \cdot \rho(x_n, x^*).$$

Therefore

$$\lim_{n\to\infty} \rho(x_{n+1}, U(x^*)) = 0,$$

i.e. $U(x^*) = \lim x_n$, and from this, by virtue of the uniqueness of the limit, (4) follows. Thus, a fixed point *exists*. It remains to verify that this point is *unique*; but this is very simple. In fact, if the operator U had another fixed point x besides x^*, then it would turn out that

$$\rho(x^*, x) = \rho(U(x^*),\ U(x)) \leqslant q \cdot \rho(x^*, x),$$

which for $\rho(x^*, x) > 0$ contradicts the condition $q < 1$. The theorem is proved completely.

Remark 1. It is clear from the very proof of the Banach theorem that *the point x^* is obtained by a method of successive approximations, starting from an arbitrary point x_0 of the space.* This remark gives a practical method of finding this point approximately. In this connection, taking the limit for fixed n and increasing m in the inequality

$$\rho(x_n, x_m) \leqslant \frac{aq^n}{1-q},$$

we obtain an estimate of the accuracy of the approximation:

$$\rho(x_n, x^*) \leqslant \frac{aq^n}{1-q}.$$

From this it is clear that the smaller q is, the more rapid the convergence of the approximation process.

Remark 2. It is interesting to observe that the condition

$$\rho(U(x'),\ U(x)) < \rho(x', x) \qquad (x' \neq x)$$

is *not sufficient* for the existence of a fixed point. For example, let E be the set of real numbers with the usual definition of distance. Set

$$U(x) = x + \frac{\pi}{2} - \arctan x.$$

Since $\arctan x < \frac{\pi}{2}$ for every x, the operator U has no fixed point. At the same time, if $x < y$, then

$$U(y) - U(x) = y - x - (\arctan y - \arctan x),$$

and by Lagrange's formula,

$$U(y) - U(x) = y - x - \frac{y-x}{1+z^2} \quad (x < z < y).$$

If we had

$$|U(y) - U(x)| \geqslant |y - x|$$

then this would mean that

$$\left| 1 - \frac{1}{1+z^2} \right| \geqslant 1,$$

but this inequality is not satisfied for any z. Therefore we always have

$$|U(y) - U(x)| < |y - x|.$$

Remark 3. Without modifying the above line of reasoning,[17] we can prove Banach's theorem in a more general form. Namely, it is sufficient to assume from the very beginning that the contraction operator U is defined only on some closed set in the complete metric space E and is such that all its values are in F. It is only necessary to choose the initial point x_0 in the process of successive approximations from F also. It is evident that the fixed point x^* will also belong to F.

We introduce some examples of the applications of Banach's theorem.

Example 1. Consider the differential equation

$$y' = f(x,y), \tag{5}$$

the right member of which is a function defined and continuous in the whole plane and which satisfies the Lipschitz condition with respect to y,

$$|f(x,y) - f(x,y_1)| \leqslant K|y - y_1|. \tag{6}$$

THEOREM. *A unique integral curve $y = \varphi(x)$ of equation* (5) *passes through any point (x_0, y_0), where the function $\varphi(x)$ satisfies equation* (5) *for all real x.*

In fact, equation (5) together with the initial condition $y(x_0) = y_0$ is completely equivalent to the integral equation

$$\varphi(x) = y_0 + \int_{x_0}^{x} f[t, \varphi(t)]\, dt. \tag{7}$$

Take any $\delta > 0$ satisfying the single condition $K\delta > 1$, and let C be the space of continuous functions defined and continuous on $[x_0 - \delta, x_0 + \delta]$. Define an operator U in C by setting

$$U(\varphi) = \psi(x)$$

for every $\varphi(x) \in C$, where

$$\psi(x) = y_0 + \int_{x_0}^{x} f[t, \varphi(t)]\, dt.$$

U is a contraction operator. In fact, if $\psi_1(x) = U(\varphi_1)$, then

$$\|\psi_1 - \psi\| = \max_{|x - x_0| < \delta} \left| \int_{x_0}^{x} \{f[t, \varphi_1(t)] - f[t, \varphi(t)]\}\, dt \right|.$$

But

$$|f[t, \varphi_1(t)] - f[t, \varphi(t)]| \leqslant K|\varphi_1(t) - \varphi(t)| \leqslant K\|\varphi_1 - \varphi\|.$$

Therefore,

$$\|\psi_1 - \psi\| \leqslant \max_{|x - x_0| < \delta} \left| \int_{x_0}^{x} K\|\varphi_1 - \varphi\|\, dt \right| = K\delta \cdot \|\varphi_1 - \varphi\|.$$

[17] Instead of repeating the same line of reasoning, one can resort to the fact that a closed set in a complete metric space is itself a complete metric space.

By Banach's principle, the operator U has a unique fixed point. Let it be $\varphi(x)$. Then $y=\varphi(x)$ is an integral curve which passes through (x_0, y_0). In this connection, the function is momentarily defined only on the segment $[x_0-\delta, x_0+\delta]$. But, setting $x_1=x_0+\delta, y_1=\varphi(x_1)$ and taking (x_1, y_1) for point of departure, we can draw an integral curve $y=\varphi_1(x)$ through it, where the function $\varphi_1(x)$ will be defined on the segment $[x_1-\delta, x_1+\delta]$ (it is important that δ be the same one as that used above!). The segments $[x_0-\delta, x_0+\delta]$ and $[x_1-\delta, x_1+\delta]$ intersect in the segment $[x_0, x_1]$. On the latter segment, both functions $\varphi(x)$ and $\varphi_1(x)$ satisfy the differential equation (5) and the condition $y(x_1)=y_1$. This means, by what was proved above, that they must coincide. Thus, by introducing the function $\varphi_1(x)$, we have extended the definition of $\varphi(x)$ from the segment $[x_0-\delta, x_0+\delta]$ to a longer segment $[x_0-\delta, x_0+2\delta]$.

Continuing this process, we convince ourselves of the possibility of extending the definition of $\varphi(x)$ to the entire interval $(-\infty, +\infty)$.

We note that the Banach principle enables one to actually construct the function $\varphi(x)$ on the segment $[x_0-\delta, x_0+\delta]$. Namely, taking an arbitrary function $\varphi_0(x)$ which is defined and continuous on this segment, and setting

$$\varphi_{n+1}(x)=y_0+\int_{x_0}^{x} f[t, \varphi_n(t)]\,dt,$$

we have the estimate

$$\max|\varphi_n(x)-\varphi(x)| \leqslant \frac{aK^n\delta^n}{1-K\delta} \qquad (x_0-\delta \leqslant x \leqslant x_0+\delta).$$

This estimate shows that the rapidity of convergence of the successive approximation process will be greater the smaller we take δ.

We have succeeded in proving the existence of a function $\varphi(x)$, satisfying equation (5) and *defined on the entire x-axis*. Such a remarkable result was obtained because very restrictive conditions were imposed upon the right member of equation (5). They are not fulfilled even for such a simple equation as, for instance, $y'=2xy$, although it has the everywhere defined solution $y=Ce^{x^2}$. It is natural to try to weaken these conditions.

Let the right member of the equation (5) be defined and continuous only in the rectangle $R(x_0-A \leqslant x \leqslant x_0+A,\ y_0-B \leqslant y \leqslant y_0+B)$ and let it satisfy the Lipschitz condition (6). Choose $\delta>0$ satisfying the condition

$$\delta < \min\left\{A, \frac{B}{M}, \frac{1}{K}\right\},$$

where $M=\max|f(x,y)|$. Then there exists a unique function $\varphi(x)$, defined on the segment $[x_0-\delta, x_0+\delta]$, satisfying equation (5) and the initial condition $\varphi(x_0)=y_0$.

In fact, let C be the space of all continuous functions defined on $[x_0-\delta, x_0+\delta]$ and let F be the set of those functions in C which satisfy the inequality

$$|\varphi(x)-y_0| \leqslant B.$$

The set F is closed in the space C. We define the operator U on F by setting

$$U(\varphi)=\psi(x)=y_0+\int_{x_0}^{x} f[t, \varphi(t)]\,dt \qquad (|x-x_0| \leqslant \delta).$$

The operator U does not take us out of the set F. In fact, for $x \in [x_0 - \delta, x_0 + \delta]$ we have

$$|\psi(x) - y_0| \leqslant \left| \int_{x_0}^{x} M\,dx \right| \leqslant M\delta \leqslant B.$$

Furthermore, as above, the operator U is a contraction operator and we can apply Remark 3 to Banach's theorem.

Note that if we do not use Banach's principle, but apply the method of successive approximations to equation (7) directly, beginning the process with $\varphi_0(x) = y_0$, we can omit the condition $K\delta < 1$.

Example 2. Consider the linear integral equation

$$\varphi(x) = f(x) + \lambda \int_a^b K(x,t)\,\varphi(t)\,dt. \tag{8}$$

Here, $f(x)$ and $K(x,t)$ are given functions which are defined and continuous on $a \leqslant x \leqslant b$ and $(a \leqslant x \leqslant b, a \leqslant t \leqslant b)$, respectively. These functions are called the *free member* and the *kernel*, respectively, of equation (8). The multiplier λ occurring before the integral sign is a numerical parameter and $\varphi(x)$ is the unknown function sought.

Having equation (8), it is possible to introduce the operator U which maps the space C of continuous functions on $[a,b]$ into itself by the formula $U(\varphi) = \psi(x)$, where

$$\psi(x) = f(x) + \lambda \int_a^b K(x,t)\,\varphi(t)\,dt.$$

Introducing the notation $M = \max |K(x,t)|$, we obtain for $U(\varphi) = \psi(x)$, $U(\varphi_1) = \psi_1(x)$, that

$$|\psi_1(x) - \psi(x)| \leqslant |\lambda| \cdot M \cdot \int_a^b |\varphi_1(t) - \varphi(t)|\,dt \leqslant |\lambda| \cdot M(b-a) \cdot \|\varphi_1 = \varphi\|.$$

Hence U is a contraction operator under the condition that

$$|\lambda| < \frac{1}{M(b-a)}. \tag{9}$$

We have thus proved the following theorem.

THEOREM. *Under condition* (9), *equation* (8) *has a unique continuous solution* $\varphi(x)$. *This solution can be obtained by the method of successive approximations by taking an arbitrary function* $\varphi_0(x) \in C$ *and setting*

$$\varphi_{n+1}(x) = f(x) + \lambda \int_a^b K(x,t)\,\varphi_n(t)\,dt.$$

Note, further, that

$$\max |\varphi_n(x) - \varphi(x)| \leqslant C|\lambda|^n M^n (b-a)^n,$$

where C is a constant which depends on the initial function $\varphi_0(x)$.

Example 3. In conclusion, we consider the problem of solving an infinite system

of linear equations. An infinite system of linear equations is a system consisting of a denumerable number of equations

$$\left.\begin{array}{l} a_{1,1}x_1 + a_{1,2}x_2 + a_{1,3}x_3 + \ldots = b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + a_{2,3}x_3 + \ldots = b_2 \\ a_{3,1}x_1 + a_{3,2}x_2 + a_{3,3}x_3 + \ldots = b_3 \\ \ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots \end{array}\right\} \tag{10}$$

in which the coefficients $a_{i,k}$ and the free members b_i are given, and the numbers x_k are the unknowns sought. A *solution* of the system (10) is a sequence of values of the unknowns which upon substitution into the left members of the equations transform them into convergent series with sums equal to the right members.

If we transpose the left member of each equation to the right, add x_i to both members of the i-th equation, and set

$$-a_{i,k} = c_{i,k} \quad (k \neq i), \quad 1 - a_{i,i} = c_{i,i},$$

then the system (10) will appear in the form

$$x_i = \sum_{k=1}^{\infty} c_{i,k}\, x_k + b_i \qquad (i = 1, 2, 3, \ldots). \tag{11}$$

We shall consider the system written in this form. If

$$\sum_{k=1}^{\infty} |c_{i,k}| \leqslant q < 1, \quad |b_i| \leqslant B \qquad (i = 1, 2, 3, \ldots),$$

where the constants q and B do not depend on the choice of i, then the system (11) is said to be *completely regular*. A solution $\{x_n\}$ of (11) is said to be *bounded* if

$$|x_k| \leqslant M$$

for all k.

THEOREM. *A completely regular system has a unique bounded solution. This solution can be found by the method of successive approximations, where we can take the members of an arbitrary bounded numerical sequence for the initial values of the unknowns.*

To prove this, we consider the space m of all bounded numerical sequences and we define an operator U in this space by setting $U(x) = y$ for the sequence $x = \{x_k\} \in m$, where $y = \{y_i\}$ is a sequence of numbers,

$$y_i = \sum_{k=1}^{\infty} c_{i,k} x_k + b_i \qquad (i = 1, 2, 3, \ldots).$$

We shall first verify that we have really defined an operator. In fact, if $\|x\| = \sup\{|x_k|\}$, then

$$|y_i| \leqslant \sum_{k=1}^{\infty} |c_{i,k}| \cdot \|x\| + |b_i| \leqslant q\|x\| + B. \tag{12}$$

This shows that our operator assigns a sequence of numbers y_i to each point $x \in m$. Furthermore, these numbers are bounded. This means that the operator maps the space m into itself. In this connection, it is useful to note that, in virtue of (12), we have

$$\|U(x)\| \leqslant B + q\|x\|. \tag{13}$$

We shall further show that U is a contraction operator. If $\{x_k\} = x$, $\{x'_k\} = x'$, $U(x) = \{y_i\}$, $U(x') = \{y'_i\}$, then

$$y_i' - y_i = \sum_{k=1}^{\infty} c_{i,k}(x_k' - x_k),$$

from which it follows that

$$|y_i' - y_i| \leqslant q \| x' - x \| \text{ and}$$

$$\| U(x') - U(x) \| \leqslant q \cdot \| x' - x \|,$$

and this we know means that U is a contraction operator.

Now the theorem follows directly from the Banach principle.[18]

Remark 1. It is not difficult to give an estimate for the solution of the system. To do this, we take the point $x^{(0)} = \Theta = (0, 0, 0, \ldots)$ for the initial point in the successive approximations process and set $x^{(n+1)} = U(x^{(n)})$. According to (13), we will have

$$\| x^{(1)} \| \leqslant B, \| x^{(2)} \| \leqslant B + qB, \| x^{(3)} \| \leqslant B + qB + q^2 B$$

and, in general,

$$\| x^{(n+1)} \| \leqslant B + Bq + Bq^2 + \ldots + Bq^n,$$

which is easily verified by induction. But then

$$\| x \| \leqslant \frac{B}{1-q}.$$

Furthermore, as always,

$$\| x^{(n)} - x \| \leqslant \frac{\| x^{(1)} - x^{(0)} \|}{1-q} \cdot q^n.$$

This means that we will have

$$\| x^{(n)} - x \| \leqslant \frac{B}{1-q} \cdot q^n$$

for the indicated choice of initial point.

Remark 2. It was already shown in the theorem that a completely regular system has a unique bounded solution. But it can have unbounded solutions besides. For example, the completely regular system

$$\begin{aligned} x_1 &= \tfrac{1}{2} x_2 \\ x_2 &= \tfrac{1}{2} x_3 \\ x_3 &= \tfrac{1}{2} x_4 \\ &\ldots \\ &\ldots \end{aligned}$$

has an infinite set of solutions $(c, 2c, 4c, 8c, \ldots)$, where c is an arbitrary constant (this is the *general form* of the solution of the indicated system). All these solutions, except the solution $(0, 0, 0, \ldots)$, are not bounded.

Remark 3. The problem sometimes arises of finding a solution $(x_1, x_2, x_3, \ldots)$ of an infinite system which is in l_p $(p \geqslant 1)$. We stop to consider this problem for the case $p = 2$.

Let the system (11) satisfy the conditions

---

[18] And the completeness of the space m.

$$\sum_{i=1}^{\infty}\sum_{k=1}^{\infty} c_{i,k}^2 = q^2 < 1, \qquad \sum_{i=1}^{\infty} b^2 < +\infty. \tag{14}$$

If $x = \{x_k\}$ is a point in l_2, then the series

$$\sum_{k=1}^{\infty} c_{i,k}x_k$$

converges and, by the Schwarz-Bunyakovsky inequality, its sum z_i will satisfy the inequality

$$z_i^2 \leqslant \left(\sum_{k=1}^{\infty} c_{i,k}^2\right) \|x\|^2.$$

But then the sequence $(z_1, z_2, z_3, \ldots)$, and with it, the sequence $(y_1, y_2, y_3, \ldots)$ where

$$y_i = \sum_{k=1}^{\infty} c_{i,k}x_k + b_i = z_i + b_i,$$

lies in l_2. This means that the operator U, which maps $\{x_k\}$ into $\{y_i\}$, is defined on l_2 and has range of variation in l_2. Furthermore, if $x = \{x_k\}$ and $x' = \{x_k'\}$ are two points in l_2, and $y = \{y_i\} = U(x)$, $y' = \{y_i'\} = U(x')$, then

$$y_i' - y_i = \sum_{k=1}^{\infty} c_{i,k}(x_k' - x_k),$$

from which it follows that

$$(y_i' - y_i)^2 \leqslant \left(\sum_{k=1}^{\infty} c_{i,k}^2\right) \cdot \|x' - x\|^2$$

and

$$\|y' - y\| \leqslant q\|x' - x\|.$$

This means that U is a contraction operator. From this it follows that, under conditions (14), the system (11) has a unique solution in l_2.

Analogously, for arbitrary $p > 1$, the conditions

$$\sum_{i=1}^{\infty}\left(\sum_{k=1}^{\infty} |c_{i,k}|^{\frac{p}{p-1}}\right)^{p-1} < 1, \qquad \sum_{i=1}^{\infty} |b_i|^p < +\infty$$

guarantee the existence and uniqueness of the solution of the system (11) in l_p. For $p = 1$, the conditions for existence and uniqueness are:

$$\sum_{i=1}^{\infty} \alpha_i < 1, \sum_{i=1}^{\infty} |b_i| < +\infty.$$

Here, for brevity, we have set $\alpha_i = \sup\{|c_{i,k}|\}$ $(k = 1, 2, 3, \ldots)$. The proof is left to the reader.

APPENDICES

1. ARC LENGTH OF A CURVE

In order to elucidate the geometric meaning of the concept of a function of finite variation, we shall consider the problem of the rectification of the arc length of a curve. For the sake of simplicity, we shall restrict ourselves to curves whose equations have the form

$$y=f(x) \qquad (a \leqslant x \leqslant b), \tag{1}$$

where $f(x)$ is continuous on $[a, b]$. As is known, the limit s of the length of the inscribed polygonal arc as the longest of its links tends to zero is called the length of curve (1). If the abscissas of the vertices of the polygonal arc are $a=x_0<x_1<\ldots<x_n=b$, then

$$s=\lim_{\lambda\to 0}\sum_{k=0}^{n-1}\sqrt{(x_{k+1}-x_k)^2+[f(x_{k+1})-f(x_k)]^2}, \tag{2}$$

where [1] $\lambda=\max(x_{k+1}-x_k)$. If $s<+\infty$, then the curve is said to be *rectifiable*.

LEMMA. *Let the sum*

$$\sigma=\sum_{k=0}^{n-1}\sqrt{(x_{k+1}-x_k)^2+[f(x_{k+1})-f(x_k)]^2} \tag{3}$$

correspond to the subdivision $a=x_0<x_1<\ldots<x_n=b$. *If a new subdivision point* $\bar{x}\in(x_i, x_{i+1})$ *is added and if we denote the sum obtained after the addition of* $\bar{x}$ *by* $\bar{\sigma}$, *then it turns out that*

$$\sigma \leqslant \bar{\sigma} \leqslant \sigma+2[x_{i+1}-x_i+\omega_i],$$

where ω_i *is the oscillation of* $f(x)$ *on* $[x_i, x_{i+1}]$.

Proof. The inequality $\sigma \leqslant \bar{\sigma}$ is clear. On the other hand, the difference $\bar{\sigma}-\sigma$ is less than or equal to the sum of the two new terms each of which is not greater than [2] $(x_{i+1}-x_i)+\omega_i$.

THEOREM 1. *Every curve* (1) *has a finite or infinite length* s. *This length is equal to the least upper bound of the set of sums* (3), *corresponding to all possible subdivisions of* $[a, b]$.

Proof. Suppose L is the least upper bound referred to in the theorem. We denote any number less than L by L_0 and let $L_0<M<L$. By the definition of the least upper bound, we can find a subdivision

$$x_0^*=a<x_1^*<\ldots<x_m^*=b, \tag{4}$$

such that the sum σ^* which corresponds to it satisfies the relation

---

[1] Strictly speaking, the smallness of the links should have been characterized by the quantity

$$\mu=\max\sqrt{(x_{k+1}-x_k)^2+[f(x_{k+1})-f(x_k)]^2}$$

rather than by the quantity λ, but it is easily shown that the relations $\lambda\to 0$ and $\mu\to 0$ are equivalent.

[2] Inasmuch as $\sqrt{a^2+b^2}\leqslant|a|+|b|$.

$$\sigma^* > M.$$

In view of the fact that $f(x)$ is continuous, an $\eta > 0$ can be found such that

$$|f(x'') - f(x')| < \frac{M - L_0}{4m} \text{ for } |x'' - x'| < \eta.$$

We denote the smallest of the distances between points (4) by d and let

$$\delta = \min\left\{d, \frac{M - L_0}{4m}, \eta\right\}.$$

We consider any subdivision

$$x_0 = a < x_1 < \ldots < x_n = b, \tag{5}$$

which is subject to the condition that $\lambda = \max(x_{k+1} - x_k) < \delta$, and we denote the sum (3) corresponding to it by σ.

In order to estimate σ, we introduce another—composite—method of subdivision of $[a, b]$, formed from the combination of subdivisions (4) and (5), and let the sum (3) corresponding to this "composite" subdivision be $\bar{\sigma}$. Since the composite subdivision is obtained from (4) by the addition of new points, we have, according to the lemma, that

$$\bar{\sigma} \geqslant \sigma^* > M. \tag{6}$$

On the other hand, the composite subdivision is obtained by means of an $(m-1)$-fold addition of one subdivision point each to the points of (5). All the points x_k^* lie in various segments $[x_i, x_{i+1}]$ and the occurrence of each of them increases the sum (3) by not more than $\frac{M - L_0}{m}$. It follows from this that

$$\bar{\sigma} \leqslant \sigma + (m-1)\frac{M - L_0}{m} < \sigma + (M - L_0).$$

Now this and (6) imply that

$$\sigma > L_0.$$

Thus, to arbitrary $L_0 < L$ there corresponds a δ such that $\lambda < \delta$ implies that $\sigma > L_0$. Since we always have $\sigma \leqslant L$, we have proved that L is the limit of (2).

THEOREM 2 (C. JORDAN). *A necessary and sufficient condition that the curve* (1) *be rectifiable is that $f(x)$ be of finite variation.*

Proof. Since, for an arbitrary subdivision, the sum

$$V = \sum_{k=0}^{n-1} |f(x_{k+1}) - f(x_k)|$$

is less than or equal to sum (3) which corresponds to this subdivision, the necessity of the condition is evident.

On the other hand,

$$\sigma = \sum_{k=0}^{n-1} \sqrt{(x_{k+1} - x_k)^2 + [f(x_{k+1}) - f(x_k)]^2} \leqslant \sum_{k=0}^{n-1} \{[x_{k+1} - x_k] + |f(x_{k+1}) - f(x_k)|\},$$

from which it follows that

$$\sigma \leqslant b - a + \overset{b}{\underset{a}{V}}(f),$$

and the condition of the theorem is then sufficient.

THEOREM 3. *If the function $f(x)$ is absolutely continuous on $[a, b]$, then the length s of the curve can be computed according to the formula*

$$s = \int_a^b \sqrt{1 + f'^2(x)}\, dx. \tag{7}$$

Proof. Since $\sqrt{1 + f'^2(x)} \leqslant 1 + |f'(x)|$, integral (7) is finite. We shall first show that

$$\sqrt{(b-a)^2 + [f(b) - f(a)]^2} \leqslant \int_a^b \sqrt{1 + f'^2(x)}\, dx. \tag{8}$$

To this end, we note that the left member of (8) is

$$\sqrt{(b-a)^2 + \left[\int_a^b f'(x)\, dx\right]^2}.$$

We set

$$b - a = r\cos\theta, \qquad \int_a^b f'(x)\, dx = r\sin\theta.$$

Then

$$\sqrt{(b-a)^2 + \left[\int_a^b f'(x)\, dx\right]^2} = r = \int_a^b [\cos\theta + f'(x)\sin\theta]\, dx.$$

It remains to note that[3]

$$|\cos\theta + f'(x)\sin\theta| \leqslant \sqrt{1 + f'^2(x)}.$$

If we subdivide $[a, b]$ by means of the points x_k and apply (8) to each of the segments $[x_k, x_{k+1}]$ we easily arrive at the estimate

$$s \leqslant \int_a^b \sqrt{1 + f'^2(x)}\, dx. \tag{9}$$

We shall now prove the reverse inequality. To this end, we subdivide $[a, b]$ into n equal parts by means of the points $x_k^{(n)} = a + \frac{k}{n}(b-a)\,(k = 0, 1, \ldots, n)$ and introduce the function $f_n(x)$ by setting

$$f_n(x) = \frac{f(x_{k+1}^{(n)}) - f(x_k^{(n)})}{x_{k+1}^{(n)} - x_k^{(n)}}, \text{ if } x_k^{(n)} < x < x_{k+1}^{(n)}.$$

We set $f_n(x_k^{(n)}) = 0$. We have, almost everywhere, that[4]

$$\lim_{n\to\infty} f_n(x) = f'(x).$$

This means, according to the Fatou theorem, that

---

[3] In virtue of the Schwarz-Bunyakovsky inequality.

[4] See the footnote to the proof of Theorem 7, §4, Chapt. IX.

$$\int_a^b \sqrt{1+f'^2(x)}\,dx \leqslant \sup\left\{\int_a^b \sqrt{1+f_n^2(x)}\,dx\right\}.$$

But

$$\int_a^b \sqrt{1+f_n^2(x)}\,dx = \sum_{k=0}^{n-1} \sqrt{(x_{k+1}^{(n)}-x_k^{(n)})^2+[f(x_{k+1}^{(n)})-f(x_k^{(n)})]^2} \leqslant s,$$

from which,[5] in combination with (9), (7) follows.

Remark. It can be proved that the presence of equality (7) (under the condition that $s<+\infty$) is also *sufficient* that $f(x)$ be absolutely continuous.

II. STEINHAUS'S EXAMPLE

In §4, Chapt. X, we saw that the coefficients of a trigonometric series which converges on a set of positive measure tend to zero. It is interesting to note that the converse of this proposition is not true.

For example,[6] the series

$$\sum_{k=3}^{\infty} \frac{\cos k(x-\ln\ln k)}{\ln k}, \tag{1}$$

whose coefficients obviously tend to zero, *diverges everywhere.*

We shall use the following notation in the proof of this assertion:[7]

$$u_k=[\ln k],\quad v_k=\ln\ln k,\quad \Delta_n=[v_n, v_{n+1}),$$

$$G_n(x)=\sum_{k=n+1}^{n+u_n} \frac{\cos k(x-v_k)}{\ln k},$$

$$G_n=\sum_{k=n+1}^{n+u_n} \frac{1}{\ln k}.$$

The smallest term in the sum G_n is the last one and hence

$$G_n > \frac{u_n}{\ln(n+u_n)}.$$

But the right member of this inequality tends to 1 as n increases and, hence, for $n>n_0$, we have

---

[5] We see that the inequality

$$\int_a^b \sqrt{1+f'^2(x)}\,dx \leqslant s$$

is also valid without the assumption that $f(x)$ is absolutely continuous. It is sufficient that $f(x)$ be a function of finite variation.

[6] This example is due to the Polish mathematician H. Steinhaus. A. N. Kolmogorov constructed the Fourier series of a summable function which diverges everywhere.

[7] $[x]$ is the entire part of x.

$$G_n > 0.9. \tag{2}$$

On the other hand,

$$G_n - G_n(x) < \frac{1}{\ln n} \sum_{k=n+1}^{n+u_n} [1 - \cos k(x - v_k)]$$

and since

$$1 - \cos\alpha = 2\sin^2\frac{\alpha}{2} \leqslant \frac{\alpha^2}{2},$$

we have

$$G_n - G_n(x) < \frac{1}{2\ln n} \sum_{k=n+1}^{n+u_n} k^2(x - v_k)^2.$$

We note that by the Lagrange formula for $\ln\ln x$,

$$v_{n+u_n} - v_n = \frac{u_n}{(n+\theta u_n)\ln(n+\theta u_n)} < \frac{u_n}{n\ln n} < \frac{1}{n}.$$

If $x \in \Delta_n$, then[8] $v_n \leqslant x < v_{n+u_n}$. Hence, for $n < k \leqslant n+u_n$ we have $|x - v_k| < v_{n+u_n} - v_n < \frac{1}{n}$ and

$$\sum_{k=n+1}^{n+u_n} k^2(x - v_k)^2 < \frac{(n+u_n)^2 u_n}{n^2}.$$

From this and (3) it follows that, for $x \in \Delta_n$,

$$G_n - G_n(x) < \frac{(n+u_n)^2 u_n}{2n^2 \ln n}.$$

The right member of this inequality tends to 0.5 as n increases. Hence, for $n > n_1$ it is less than 0.6, from which in combination with (2) we see that for $n > \max(n_0, n_1)$ and for all x in Δ_n, we have

$$G_n(x) > 0.3.$$

Now the proof can be easily completed. Namely, we fix an arbitrary x_0 and then set

$$x_i = x_0 + 2\pi i \qquad (i = 1, 2, 3, \ldots).$$

Each of the points x_i (or at least beginning with some one of them) lies in some interval Δ_{n_i}, where the numbers n_i increase indefinitely with i. In view of the fact that the sums $G_n(x)$ have period 2π, we have

$$G_{n_i}(x_0) = G_{n_i}(x_i) > 0.3.$$

This proves that series (1) diverges at the point x_0 since $G_n(x)$ is the difference of the partial sums S_{n+u_n} and S_n of this series.

---

[8]Here we have taken into account that $n \geqslant 3$, from which it follows that $u_n \geqslant 1$.

III. CERTAIN SUPPLEMENTARY INFORMATION ABOUT CONVEX FUNCTIONS

Here, we shall furnish additional material to §5, Chapt. X.

LEMMA. *If $f(x)$ is a function defined on (a,b) which is convex downward and $c\in(a,b)$, then for $x\neq c$ the ratio*

$$\frac{f(x)-f(c)}{x-c} \tag{1}$$

is an increasing function of x.

Proof. Suppose $c<x<y<b$. Then[9]

$$f(x)=f\left[\frac{y-x}{y-c}c+\frac{x-c}{y-c}y\right]\leqslant\frac{y-x}{y-c}f(c)+\frac{x-c}{y-c}f(y).$$

It follows that

$$\frac{f(x)-f(c)}{x-c}\leqslant\frac{f(y)-f(c)}{y-c}. \tag{2}$$

We prove (2) analogously for the cases $a<x<y<c$ and $a<x<c<y<b$.

THEOREM 1. *Under the conditions of the lemma, the function $f(x)$ satisfies the Lipschitz condition on every segment $[p,q]\subset(a,b)$.*

Proof. Suppose x and y are two points in $[p,q]$ and let $m=\frac{q+b}{2}$. According to the lemma,

$$\frac{f(y)-f(x)}{y-x}\leqslant\frac{f(m)-f(x)}{m-x}\leqslant\frac{|f(m)|+|f(x)|}{m-q}. \tag{3}$$

Since the function $f(x)$ is bounded on $[p,q]$, the right member of (3) is bounded above, which fact leads to the inequality

$$\frac{f(y)-f(x)}{y-x}\leqslant K.$$

It is proved analogously that

$$\frac{f(y)-f(x)}{y-x}\geqslant L.$$

The remainder of the proof is clear.

COROLLARY. *Under the conditions of the lemma, we have*

$$f(x)=f(c)+\int_c^x\varphi(t)\,dt, \tag{4}$$

where c is an arbitrary point in (a,b), and $\varphi(t)$ is a measurable function which is bounded on every $[p,q]\subset(a,b)$.

THEOREM 2. *The class of functions which are convex downward on the interval (a,b) coincides with the class of indefinite integrals of functions which are increasing on (a,b) and bounded on every $[p,q]\subset(a,b)$.*

---

[9]On the basis of inequality (16), §5, Chapt. X. Since the function $f(x)$ is continuous on (a,b), one can apply (16).

Proof. The fact that every such indefinite integral is a function which is convex downward was proved in Chapt. X. To prove the converse, we shall show that the function $\varphi(t)$ appearing in (4) can be assumed increasing. Suppose $x < y$ are two points in (a, b) and that $h > 0$ is so small that $x + h < y$ and $y + h < b$.

According to the lemma, the ratio $\dfrac{f(x+h)-f(x)}{h}$ does not decrease when y is replaced by $x + h$, from which it follows that

$$\frac{f(x+h)-f(x)}{h} \leqslant \frac{f(y)-f(x)}{y-x} = \frac{f(x)-f(y)}{x-y}.$$

If we replace x by $y + h$, the inequality is only strengthened. Hence,

$$\frac{f(x+h)-f(x)}{h} \leqslant \frac{f(y+h)-f(y)}{h}.$$

Assuming that the derivatives $f'(x)$ and $f'(y)$ exist at the points x and y, we find that $f'(x) \leqslant f'(y)$.

Denoting the set of all points at which $f'(x)$ exists by E and setting

$$\varphi(t) = \sup\{f'(x)\} \qquad (x \leqslant t, x \in E)$$

for all $t \in (a, b)$, we complete the proof.

Suppose the convex function $M(x)$, defined on $[0, +\infty)$, is representable in the form

$$M(x) = \int_0^x m(t)\,dt,$$

where $m(t)$ is a continuous, strictly increasing function such that $m(0) = 0, m(+\infty) = +\infty$. Then the function $z = m(t)$ has an inverse $t = n(z)$, defined on the segment $0 \leqslant z < +\infty$, which is also continuous and strictly increasing, with $n(0) = 0$, $n(+\infty) = +\infty$. We set

$$N(x) = \int_0^x n(z)\,dz.$$

Like $M(x)$, the function $N(x)$ is increasing and convex downward on $[0, +\infty)$.

THEOREM 3. *Using the notation introduced above, we have for arbitrary $a \geqslant 0$ and $b \geqslant 0$ that the so-called "Young inequality" is valid:*

$$ab \leqslant M(a) + N(b). \tag{5}$$

Proof.[10] It is easily seen that the integral

$$N(b) = \int_0^b n(z)\,dz$$

can be represented[11] in the form of a Stieltjes integral

[10]Geometrically, (5) is almost obvious.

[11]If $m(t)$ is continuous and strictly increasing on $[p, q]$, and $f(z)$ is continuous on $[m(p), m(q)]$, then

$$(S)\int_p^q f[m(t)]\,dm(t) = (R)\int_{m(p)}^{m(q)} f(z)\,dz.$$

This follows directly from the consideration of the corresponding integral sums.

$$N(b) = \int_0^{n(b)} t\, dm(t).$$

Integrating by parts and taking into consideration that $m[n(b)] = b$, we find that

$$N(b) = bn(b) - M[n(b)].$$

It follows that

$$ab = bn(b) - [n(b) - a]b = M[n(b)] + N(b) - [n(b) - a]b,$$

or

$$ab = M(a) + N(b) + \left\{ \int_a^{n(b)} m(t)\, dt - [n(b) - a]b \right\}.$$

It remains to show that

$$\int_a^{n(b)} m(t)\, dt \leqslant [n(b) - a]b. \tag{6}$$

We shall assume that $a \leqslant n(b)$ [the case $a > n(b)$ is treated analogously]. Since $m(t)$ increases, we have that for $t \leqslant n(b), m(t) \leqslant m[n(b)] = b$, from which (6) follows.

DEFINITION. The functions $M(x)$ and $N(x)$, defined on $[0, +\infty)$, are said to be *mutually complementary in the Young sense* if (5) holds for arbitrary $a \geqslant 0$ and $b \geqslant 0$.

THEOREM 4 (Z. BIRNBAUM—W. ORLICZ). *If $M(x)$ is continuous on $[0, +\infty)$, $M(0) = 0, M(x) \geqslant 0$ for $x > 0$ and*

$$\lim_{x \to +\infty} \frac{M(x)}{x} = +\infty, \tag{7}$$

then there exists a non-negative function $N(x)$ which is complementary to $M(x)$ in the Young sense, continuous on $[0, +\infty)$, increasing and convex downward.

Proof. We assign the function

$$\varphi_x(y) = xy - M(y)$$

to each $x \geqslant 0$.

Clearly, $\varphi_x(0) = 0$. In virtue of (7), for sufficiently large y, we shall have $\varphi_x(y) < 0$. Suppose this inequality holds for $y \geqslant y_0$. On the segment $[0, y_0]$ the continuous function $\varphi_x(y)$ has a maximal value which obviously will also be its maximal value on the entire half-axis $[0, +\infty)$. We denote this value by $N(x)$:

$$N(x) = \max_{y \geqslant 0} [xy - M(y)]. \tag{8}$$

It is easily seen that $N(0) = 0$. We shall show that $N(x)$ satisfies all the requirements of the theorem.

Since $N(x) \geqslant \varphi_x(0) = 0, N(x)$ is non-negative. If x is held fixed, then for arbitrary y we shall have $xy - M(y) \leqslant N(x)$, from which it is clear that $N(x)$ is complementary to $M(x)$ in the Young sense. Further, for arbitrary $x_1 \geqslant 0$ and $x_2 \geqslant 0$, we shall have

$$N\left(\frac{x_1 + x_2}{2}\right) = \max \left[\frac{x_1 + x_2}{2} y - M(y)\right].$$

But

$$\frac{x_1 + x_2}{2} y - M(y) = \frac{x_1 y - M(y)}{2} + \frac{x_2 y - M(y)}{2} \leqslant \frac{N(x_1) + N(x_2)}{2}.$$

This proves that $N(x)$ is convex and hence that it is continuous everywhere except at $x=0$. We shall prove that the function $N(x)$ is continuous also for $x=0$.

In virtue of (7), the set of values y for which $M(y) \leqslant y$ is bounded above; this set is evidently non-void and closed. Suppose p is its rightmost point. If $0< x \leqslant 1$, then for all $y>p$, we shall have

$$\varphi_x(y) = xy - M(y) < xy - y \leqslant 0.$$

Hence, the maximal value of $\varphi_x(y)$ is assumed at a point y_0 such that $y_0 \leqslant p$. Thus,

$$N(x) = xy_0 - M(y_0) \leqslant xy_0 \leqslant px.$$

It follows from this that the function $N(x)$ is continuous at $x=0$.

It remains to show that $N(x)$ increases. We take the points x_1, x_2: $0< x_1< x_2$. Then

$$N(x_1) = N\left[\frac{x_2 - x_1}{x_2}\cdot 0 + \frac{x_1}{x_2}x_2\right] \leqslant \frac{x_2 - x_1}{x_2}N(0) + \frac{x_1}{x_2}N(x_2),$$

and since $N(0)=0$, we have

$$N(x_1) \leqslant N(x_2).$$

This proves the theorem completely.

Remarks. 1) The function $N(x)$ constructed with the aid of formula (8) is such that

$$\lim_{x\to+\infty} \frac{N(x)}{x} = +\infty.$$

In fact, taking arbitrary n and ε, we can find an x_0 such that $x > x_0$ implies that $M(n+\varepsilon) < \varepsilon x$. For these x, we have from (5) that

$$(n+\varepsilon)x \leqslant M(n+\varepsilon) + N(x) < \varepsilon x + N(x),$$

from which it follows that $N(x) > nx$.

2) Theorem 4 generalizes Theorem 3, because (7) is also satisfied under the conditions of the latter theorem. In fact, taking an arbitrary $A>0$, an a can be found such that for $t \geqslant a$ we have $m(t) \geqslant A$. If $x > a$, then

$$\frac{M(x)}{x} > \frac{M(a) + A(x-a)}{x}. \tag{9}$$

As x increases, the right member of (9) tends to A. This means that

$$\lim_{x\to+\infty} \frac{M(x)}{x} \geqslant A,$$

which proves (7).

3) It is not assumed in Theorem 4 that $M(x)$ is convex and we do not require[12] that

$$\lim_{x\to 0} \frac{M(x)}{x} = 0. \tag{10}$$

On the contrary, under the conditions of Theorem 3 we will have $M(x) \leqslant xm(x)$,

[12]For example, Theorem 4 is applicable to the function

$$M(x) = \begin{cases} \sqrt{x} & \text{for } 0 \leqslant x \leqslant 1, \\ x^2 & \text{for } 1 \leqslant x < +\infty. \end{cases}$$

and (10) is satisfied. If the condition $M(x) \geqslant 0$ of Theorem 4 is strengthened by requiring that $M(x) > 0$ for $x > 0$, then we could prove that the function $N(x)$ defined by formula (8) satisfies the condition[13]

$$\lim_{x \to 0} \frac{N(x)}{x} = 0. \tag{11}$$

Thus, $N(x)$ is "better" than $M(x)$.

4) If $M(x)$ satisfies the conditions of Theorem 3, then formula (8) yields the same $N(x)$ as Theorem 3. In fact, in this case, we have

$$\frac{d\varphi_x(y)}{dy} = x - m(y).$$

A solution of this equation is $y = n(x)$. Hence

$$N(x) = xn(x) - \int_0^{n(x)} m(t)\,dt.$$

Setting $m(t) = z$ in this integral and integrating the integral thus obtained by parts, we obtained the required result.

5) Applying Theorem 3 to $m(t) = t^{p-1} (p > 1)$, we obtain the Hölder inequality

$$ab \leqslant \frac{a^p}{p} + \frac{b^q}{q} \left(\frac{1}{p} + \frac{1}{q} = 1\right).$$

IV. CHANGE OF VARIABLE IN THE LEBESGUE INTEGRAL

It is generally known what great significance the problem of change of variable of integration has in integral calculus. Here, we shall consider this problem for Lebesgue integrals, where we shall restrict ourselves to the case[14] when the old variable of integration x is a strictly monotonic and absolutely continuous function of the new variable t. For definiteness, we shall assume that this function is increasing. Thus, let

$$x = \varphi(t)\,[p \leqslant t \leqslant q, a = \varphi(p), b = \varphi(q)\,],$$

where $\varphi(t)$ is absolutely continuous and strictly increasing on $[p,q]$.

LEMMA 1. *If E_t is a measurable set contained in $[p,q]$, and $E_x = \varphi(E_t)$ is its image under the mapping $x = \varphi(t)$, then*[15]

$$mE_x = \int_{E_t} \varphi'(t)\,dt. \tag{1}$$

Proof. The measurability of E_x was proved in §3, Chapt. IX. For the case when $E_t = [\alpha, \beta]$, formula (1) is clear because in this case $E_x = [\varphi(\alpha), \varphi(\beta)]$ and

---

[13](11) may not hold without the additional requirement pointed out in the preceding sentence.

[14]A more extensive discussion of this problem can be found in C. de la Vallée Poussin: *Cours d'analyse infinitésimale*, Gauthiers-Villars, Paris.

[15]In order that the symbol $\varphi'(t)$ be everywhere defined, we shall agree (as in Chapt. VIII) that $\varphi'(t) = 0$ on the set (of measure 0) where $\varphi(t)$ has no derivative.

$$mE_x = \varphi(\beta) - \varphi(\alpha) = \int_\alpha^\beta \varphi'(t)\,dt.$$

The situation is analogous when $E_t = (\alpha, \beta)$. But then it is clear that (1) holds whenever E_t is an open set (in this case E_x will also be an open set). Going over to the complements, we easily verify the validity of (1) for the case when E_t is a closed set. Finally, we consider the general case when E_t is an arbitrary measurable set. Taking $\varepsilon > 0$, we can find a closed set F_x and an open set G_x such that[16]

$$F_x \subset E_x \subset G_x \subset (a, b), mF_x > mE_x - \varepsilon, mG_x < mE_x + \varepsilon.$$

Suppose F_t and G_t are the inverse images of the sets F_x and G_x. By what we proved above (clearly F_t is closed and G_t is open), we shall have

$$\int_{F_t} \varphi'(t)\,dt = mF_x, \int_{G_t} \varphi'(t)dt = mG_x.$$

Since $\varphi'(t) \geqslant 0$, we have

$$\int_{F_t} \varphi'(t)\,dt \leqslant \int_{E_t} \varphi'(t)\,dt \leqslant \int_{G_t} \varphi'(t)\,dt.$$

Thus,

$$mF_x \leqslant \int_{E_t} \varphi'(t)\,dt \leqslant mG_x$$

and, a fortiori,

$$mE_x - \varepsilon < \int_{E_t} \varphi'(t)\,dt < mE_x + \varepsilon.$$

Since ε is arbitrary, this completes the proof of the lemma.

LEMMA 2. *Suppose e_x is a measurable set of measure* 0, *contained in* $[a, b]$, *and that e_t is its inverse image. Then the subset e_t^* of those points of e_t for which the relation*[17]

$$\varphi'(t) = 0 \tag{2}$$

is not satisfied has measure 0.

Proof. We note that the measurability of the set e_t is no longer presupposed.[18] We can assume that $e_x \subset (a, b)$; we construct the sequence of open sets

$$(a, b) \supset G_x^{(1)} \supset G_x^{(2)} \supset G_x^{(3)} \supset \ldots, G_x^{(n)} \supset e_x, mG_x^{(n)} \to 0$$

and let

$$E_x = \prod_{n=1}^{\infty} G_x^{(n)}.$$

If $G_t^{(n)}$ is the inverse image of $G_x^{(n)}$, and E_t is the inverse image of E_x, then $G_t^{(n)}$ is open and

---

[16]Without loss of generality, we may assume that p and q do not lie in E_t.

[17]In other words, e_t^* is that subset of e_t where $\varphi'(t) > 0$ (which can be infinite) exists.

[18]Although the inverse function of $\varphi(t)$ exists, the inverse function does not necessarily have to be absolutely continuous.

$$E_t = \prod_{n=1}^{\infty} G_t^{(n)},$$

from which it follows that E_t is measurable. Clearly, $mE_x = 0$. Hence, by Lemma 1,

$$\int_{E_t} \varphi'(t)\,dt = 0. \tag{3}$$

We denote the set of those points in E_t at which (2) is not satisfied by E_t^*; then it follows from (3) that $mE_t^* = 0$. It remains to note that $e_t^* \subset E_t^*$, because $e_x \subset E_x$ and $e_t \subset E_t$.

THEOREM. *Suppose $f(x)$ is a summable function on*[19] $[a,b]$. *Then*

$$\int_a^b f(x)\,dx = \int_p^q f[\varphi(t)]\,\varphi'(t)\,dt. \tag{4}$$

Proof. We shall first consider the case when $f(x)$ is continuous[20] on $[a,b]$, so that the integral in the left member of (4) can be taken in the Riemann sense. We subdivide $[a,b]$ by the points

$$a = x_0 < x_1 < \ldots < x_n = b$$

and let M_k and m_k be the maximal and minimal values of $f(x)$ on $[x_k, x_{k+1}]$. If $\varphi(t_k) = x_k$, then for $t \in [t_k, t_{k+1}]$ we have

$$m_k \leqslant f[\varphi(t)] \leqslant M_k,$$

from which, in combination with the inequality $\varphi'(t) \geqslant 0$ and the relation

$$\int_{t_k}^{k+1} \varphi'(t)\,dt = x_{k+1} - x_k,$$

it follows that the integral

$$(L)\int_{t_k}^{t_{k+1}} f[\varphi(t)]\,\varphi'(t)\,dt$$

lies between the numbers $m_k(x_{k+1} - x_k)$ and $M_k(x_{k+1} - x_k)$. Hence, the integral in the right member of (4) is contained between the Darboux sums of the function $f(x)$ corresponding to the subdivision carried out above. Now (4) follows from this.

Suppose now that $f(x)$ is a measurable, bounded function, $|f(x)| \leqslant M$. A sequence of continuous functions $\{g_n(x)\}$ can be found for which we have

$$\lim_{n\to\infty} g_n(x) = f(x) \tag{5}$$

almost everywhere on $[a,b]$. We may assume that $|g_n(x)| \leqslant M$. Then (4) will be valid if f is replaced by g_n. It follows without difficulty from this that

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \int_p^q g_n[\varphi(t)]\,\varphi'(t)dt. \tag{6}$$

---

[19] We retain the notation introduced above.

[20] Then, clearly, $f[\varphi(t)]$ is also continuous on $[p,q]$.

We shall show that almost everywhere on $[p,q]$ we have

$$\lim_{n\to\infty} g_n[\varphi(t)]\varphi'(t) = f[\varphi(t)]\varphi'(t). \tag{7}$$

We denote by e_x the set of those x for which (5) is not satisfied. This set is measurable and $me_x = 0$. We denote the inverse image of the set e_x by e_t and let e_t^* be the subset of those t in e_t for which (2) is not satisfied.

If $t \bar{\in} e_t$, then $\varphi(t) \bar{\in} e_x$, and (7) is satisfied. In exactly the same way, if $t \in e_t - e_t^*$, then $\varphi'(t) = 0$, and (7) is again satisfied. Thus, the only points t for which (7) is not satisfied are the points in e_t^*, and by Lemma 2 we have $me_t^* = 0$.

The measurability of the product[21] $f[\varphi(t)]\varphi'(t)$ follows from the fact that (7) holds almost everywhere on $[p,q]$. Furthermore,

$$|g_n[\varphi(t)]\varphi'(t)| \leqslant M\varphi'(t)$$

and (7) guarantees the possibility of passage to the limit under the integral sign in (6) and this brings us to (4).

Suppose, finally, that $f(x)$ is an arbitrary summable function. It can be assumed that $f(x)$ is non-negative, because every summable function is the difference of two non-negative summable functions. We set

$$f_n(x) = \begin{cases} f(x) & \text{for } f(x) \leqslant n, \\ n & \text{for } f(x) > n. \end{cases}$$

According to what we have already proved, (4) will be true if f is replaced by f_n. It follows from this that

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \int_p^q f_n[\varphi(t)]\varphi'(t)\,dt. \tag{8}$$

But everywhere on $[p,q]$ we have[22]

$$f[\varphi(t)]\varphi'(t) = \lim_{n\to\infty} f_n[\varphi(t)]\varphi'(t)$$

and, furthermore, the product $f_n[\varphi(t)]\varphi'(t)$ increases with n. Therefore, in virtue of Theorem 10 (B. Levi) in §1, Chapt. VI, we will have

$$\int_p^q f[\varphi(t)]\varphi'(t)\,dt = \lim_{n\to\infty} \int_p^q f_n[\varphi(t)]\varphi'(t)\,dt,$$

from which, in combination with (8), (4) follows.

COROLLARY. *The following formulas are valid:*

$$\int_a^b f(x)\,dx = \int_{a-h}^{b-h} f(t+h)dt, \quad \int_a^b f(x)\,dx = k\int_{a/k}^{b/k} f(kt)\,dt \quad (k \neq 0).$$

---

[21] This fact is remarkable in this respect that the factor $f[\varphi(t)]$ is not at all required to be measurable.

[22] This assures the measurability of the product $f(\varphi)\varphi'$.

V. HAUSDORFF'S THEOREM

In Chapter III we already formulated Hausdorff's theorem according to which the "easy" problem of the theory of measure is not solvable for spaces of dimension greater than two. Here, we set forth the proof of this theorem. At the base of the proof lies the following

LEMMA. *The surface K of the unit sphere in three-dimensional space decomposes into four pairwise disjoint sets*

$$K = Q + R + S + T, \tag{1}$$

of which Q is denumerable; the sets R, S and T are pairwise congruent and at the same time R is congruent to the sum $S + T$. In this connection, rotations of the space about the center of the sphere are the isometries which superpose these congruent sets.

We shall first show that Hausdorff's theorem actually follows from the lemma; and then we shall prove the lemma itself. Assume that the easy problem of the theory of measure has been solved for three-dimensional space and let μE be the measure of the bounded set E.

Denote by K_0 the set of all points of the unit sphere with the exception of its center. Corresponding to the decomposition of the surface of the sphere by formula (1), the set K_0 also decomposes into four pairwise disjoint subsets

$$K_0 = Q_0 + R_0 + S_0 + T_0, \tag{2}$$

where Q_0 is the set of points situated on the rays connecting the center of the sphere with the points of the set Q, and R_0, S_0, T_0 have a similar interpretation.

We shall show that $\mu Q_0 = 0$. In fact, let us put the origin of a rectangular system of coordinates at the center of the sphere, choosing the Oz axis so that it goes through none of the points of the set Q. If we now introduce spherical coordinates, the position of a point on the sphere will be determined by two angles: "latitude" and "longitude." The set of longitude points of Q is denumerable as well as the set of differences of these longitudes. Therefore, we can choose an angle λ which is distinct from all these differences. If we rotate the space about the Oz axis through this angle λ, the set Q goes over into the set Q' which has no point in common with Q. Like the set Q, the set Q' is denumerable. Therefore, we can find an angle λ', distinct from the differences of the longitudes of the points of the set $Q + Q'$. Rotating the space about the Oz axis through the angle λ', we obtain from the set Q a new set Q'' which is congruent to Q and which intersects neither Q nor Q'. Repeating this process, we construct n pairwise disjoint congruent sets Q, Q', ..., $Q^{(n-1)}$. In this case,

$$\mu K_0 \geqslant \mu\left(\sum_{k=0}^{n-1} Q_0^{(k)}\right) = n\mu Q_0 \qquad (Q^{(0)} = Q)$$

from which it follows that $\mu Q_0 = 0$.

We note further that the rotation taking R into S carries R_0 into S_0 at the same time. Therefore, the sets R_0 and S_0 are congruent. In exactly the same way, R_0 is congruent to T_0 and to the sum $S_0 + T_0$.

From (2) and from the fact that $\mu Q_0 = 0$, it follows that

$$\mu K_0 = \mu R_0 + \mu S_0 + \mu T_0,$$

from which, in combination with the equalities $\mu R_0 = \mu S_0 = \mu T_0$, it follows that

$$\mu K_0 = 3\mu R_0.$$

At the same time, $\mu R_0 = \mu(S_0 + T_0) = \mu S_0 + \mu T_0 = 2\mu R_0$ and $\mu R_0 = 0$. This means that $\mu K_0 = 0$, which, however, contradicts the fact that for the unit cube E_0, $\mu E_0 = 1$ (because E_0 decomposes into a finite number of congruent cubes which are so small that they are "immersed" in K_0).

Thus, the non-solvability of the easy problem of measure for the space R_3 follows from the lemma. But then this problem is non-solvable for all R_n, where $n > 3$. In fact, let, for instance, this problem be solvable for R_4 and let μE^* be the measure of a bounded set $E^* \subset R_4$. Correspond to each bounded set E of three-dimensional space the set E^* consisting of those points (x, y, z, t) of the space R_4 for which $(x, y, z) \in E$, and $t \in [0, 1]$. It is not difficult to see that if we set, by definition, μE equal to μE^*, the easy problem of measure for R_3 would be solved and this, as we saw, is impossible. This means the problem is non-solvable also for R_4. We reason the same way for other R_n.

Thus, everything reduces to the proof of the lemma.

Consider a rectangular system of coordinates, whose origin is at the center of our sphere, where rotation from the positive x-axis to the positive y-axis occurs through an angle $+\frac{\pi}{2}$ in the clockwise sense if we observe from the positive z-axis. Through the origin of coordinates, we draw a u-axis in the zx-plane and situated so that the positive direction of the z-axis goes over into the positive direction of the u-axis under a rotation through the angle $\frac{1}{2}\theta$ $(0 < \theta < \pi)$ in the clockwise sense, if we observe from the positive y-axis.

We denote by φ the rotation of the space through 180° about the u-axis and by ψ the rotation of the space through the angle 120° about the z-axis (having in mind rotation in the clockwise sense, if we observe from the positive direction of the axis of rotation). We now form combinations of these rotations. As a result, we obtain the various rotations of the space about the origin of coordinates, where each such rotation will be written down as a symbol of the form[23]

$$\varphi^2\psi^4\varphi, \varphi\psi\varphi^4, \psi^5\varphi\psi^2\varphi^3, \tag{3}$$

and so on.

Inasmuch as

$$\varphi^2 = 1, \psi^3 = 1 \tag{4}$$

(here 1 symbolizes the "identity" of the group of rotations, i.e. position of rest), it is clear that our rotations can be written with "reduced" symbols, in which the "factors" φ^m with $m > 1$ and ψ^m with $m > 2$ are not encountered. For instance, the "reduced" expressions for the rotations (3) are

$$\psi\varphi, \varphi\psi, \psi^2\varphi\psi^2\varphi.$$

We shall now consider the group G of all rotations obtained with the aid of such reduced symbols. It might occur that some of these rotations will be geometrically identical, although they have different reduced symbols. Our immediate problem is to show that such a choice of the angle θ is possible for which this situation does not occur,

---

[23]In this connection, we write the rotation which is carried out *first* on the *left*.

i.e. rotations written with *distinct* reduced symbols will also be geometrically distinct.

To this end, we note that every equality of the form $\rho = \sigma$ can be written in the equivalent form $\rho\sigma^{-1} = 1$. Therefore, it will suffice to show that such a choice of the angle θ is possible for which none of the reduced symbols represents the identity of the group G.

Elements of the group G, distinct from 1, φ, ψ, and ψ^2, can have one of the following four forms:

$$\alpha = \varphi\psi^{m_1}\varphi\psi^{m_2}\ldots\varphi\psi^{m_n},\ \gamma = \varphi\psi^{m_1}\varphi\psi^{m_2}\ldots\psi^{m_n}\varphi,$$
$$\beta = \psi^{m_1}\varphi\psi^{m_2}\varphi\ldots\psi^{m_n}\varphi,\ \delta = \psi^{m_1}\varphi\psi^{m_2}\varphi\ldots\varphi\psi^{m_n},$$

where n is a natural number[24] and m_k equals 1 or 2.

We shall assume that we have succeeded in making such a choice of the angle θ for which no equality of the form $\alpha = 1$ is possible, and we shall show that under this assumption also none of the equalities

$$\beta = 1, \gamma = 1, \delta = 1$$

is possible.

In fact, it follows from the equality $\beta = 1$ that $\varphi\beta\varphi = 1$, and since

$$\varphi\beta\varphi = \varphi[\psi^{m_1}\varphi\ldots\psi^{m_n}\varphi]\varphi = \varphi\psi^{m_1}\varphi\ldots\psi^{m_n} = \alpha,$$

the equality $\beta = 1$ would imply that $\alpha = 1$, contrary to our assumption. Thus, $\beta \neq 1$. Further, it would follow from the equality $\gamma = 1$[25] that $\varphi\gamma\varphi = \delta = 1$. Hence, it only remains to verify that the equality $\delta = 1$ is impossible (under the assumption that the equality $\alpha = 1$ is impossible).

We divide the elements δ into two classes, putting δ into the first class if

$$m_1 + m_n \neq 3$$

and into the second class if $m_1 + m_n = 3$.

If δ is in the first class and $\delta = 1$, then

$$1 = \delta = \psi^{-m_1}\delta\psi^{m_1} = \varphi\psi^{m_2}\varphi\ldots\varphi\psi^{m_1+m_n} = \alpha,$$

and this is impossible by assumption.

If, however, δ is in the second class and $\delta = 1$, then

$$\varphi\psi^{m_n}\delta\psi^{m_1}\varphi = 1,$$

or

$$\varphi\psi^{m_n}[\psi^{m_1}\varphi\psi^{m_2}\ldots\varphi\psi^{m_n}]\psi^{m_1}\varphi = 1,$$

or

$$\psi^{m_2}\varphi\ldots\psi^{m_{n-1}} = 1.$$

But the element $\psi^{m_2}\varphi\ldots\psi^{m_{n-1}}$ is also of type δ (generally speaking, $\varphi\psi^{m_n}\delta\psi^{m_1}\varphi$ may coincide with φ, ψ and ψ^2, but since the last elements are known to be different from the identity, these possibilities disappear). Thus, if δ is in the second class and $\delta = 1$, then another element δ', of the same type as δ, can be found such that $\delta' = 1$, but δ' contains two less factors ψ^{m_k} than δ. Since, by what has already been proved, the element δ' cannot belong to the first class, there is the possibility of further reducing this number of factors ψ^{m_k}. In view of the fact that we cannot arrive at an element of the first class nor at the elements φ, ψ, ψ^2, it turns out to be possible to reduce the number of factors

[24] Where n must be > 1 for the elements δ.

[25] Or that $\psi^m = 1$ [if $\gamma = \varphi\psi^m\varphi$], but this is not true because $m = 1, 2$.

ψ^{m_k} *indefinitely*, which clearly is an absurdity because there were only a finite number of them from the very beginning.

Thus, the problem reduces to finding an angle θ for which none of the equalities $\alpha = 1$ is satisfied.

But every rotation α is the product of n rotations of the form $\varphi\psi$ and $\varphi\psi^2$. With the usual methods of analytic geometry, it is not difficult to show that as a result of the rotation $\varphi\psi$, the point $M(x,y,z)$ goes over into the point $N(x',y',z')$ where

$$\left.\begin{aligned} x' &= \frac{1}{2}x\cos\theta + \frac{\sqrt{3}}{2}y - \frac{1}{2}z\sin\theta \\ y' &= -\frac{\sqrt{3}}{2}x\cos\theta + \frac{1}{2}y + \frac{\sqrt{3}}{2}z\sin\theta \\ z' &= x\sin\theta + z\cos\theta \end{aligned}\right\}. \tag{5}$$

As a result of the rotation $\varphi\psi^2$ the point $M(x,y,z)$ goes over into the point $N(x',y',z')$, where x',y',z' are expressed by the same formulas (5), but with $\sqrt{3}$ replaced by $-\sqrt{3}$.

From formulas (5), by induction with respect to n, it is easy to obtain[26] that as a result of the rotation α, the point $M_0(0,0,1)$ goes over into the point $M_n(x_n,y_n,z_n)$, where

$$z_n = a_0\cos^n\theta + a_1\cos^{n-1}\theta + \ldots + a_{n-1}\cos\theta + a_n \,(a_0 \neq 0).$$

If the rotation α is identical with the identity of the group G, then $z_n = 1$ and, hence, $\cos\theta$ is a zero of the polynomial

$$P_\alpha(u) = a_0u^n + a_1u^{n-1} + \ldots + a_{n-1}u + a_n - 1.$$

Every rotation α has a polynomial $P_\alpha(u)$ corresponding to it. Since there is a denumerable set of rotations α, there is a denumerable set of polynomials $P_\alpha(u)$ and this means that the set of their zeros is denumerable. This means that it is possible to choose θ so that the $\cos\theta$ is different from all these zeros. For such θ, none of the α coincides with 1. In the sequel, we shall assume that the angle θ has been chosen in exactly this way.

We partition the elements of the group G into three disjoint classes A, B, C in such a way that the following two requirements are fulfilled:

I. *If* $\rho \in A$, *then* $\rho\varphi \in B + C$, *and conversely.*

II. *Each of the relations* $\rho \in A$, $\rho\psi \in B$, $\rho\psi^2 \in C$ *implies both of the others.*

---

[26] Three formulas must be proved inductively simultaneously:

$$x_n = \sin\theta\,[A_n^{(n)}\cos^{n-1}\theta + A_{n-1}^{(n)}\cos^{n-2}\theta + \ldots + A_1^{(n)}],$$
$$y_n = \sin\theta\,[B_n^{(n)}\cos^{n-1}\theta + B_{n-1}^{(n)}\cos^{n-2}\theta + \ldots + B_1^{(n)}],$$
$$z_n = C_n^{(n)}\cos^n\theta + C_{n-1}^{(n)}\cos^{n-1}\theta + \ldots + C_0^{(n)},$$

where we note, along the way, that $C_{n+1}^{(n+1)} = C_n^{(n)} - A_n^{(n)}$,

$$A_{n+1}^{(n+1)} = \frac{A_n^{(n)} - C_n^{(n)}}{2}, \text{ from which it follows that}$$

$$C_{n+1}^{(n+1)} - A_{n+1}^{(n+1)} = \frac{3}{2}(C_n^{(n)} - A_n^{(n)}) = \left(\frac{3}{2}\right)^n \cdot (C_1^{(1)} - A_1^{(1)}) = \left(\frac{3}{2}\right)^{n+1} \text{ and } C_n^{(n)} = \left(\frac{3}{2}\right)^{n-1}.$$

The very burdensome verification of the possibility of such a partition will be postponed until the end of the proof, and for the time being we shall *assume* that it has been carried out.

We agree, further, to use the following notation. If $\rho \in G$ and M is any point of the space, then we shall denote[27] by $\rho(M)$ that point into which M maps as a result of the rotation ρ. If, however, $E = \{M\}$ is a set of arbitrary points, then we shall denote by $\rho(E)$ the image of this entire set under the rotation ρ:

$$\rho(E) = \sum_{M \in E} \{\rho(M)\}.$$

Let Q be the set of those points of the surface K of the unit sphere which remain fixed for at least one rotation, distinct from 1, of the group G. Since the entire group is denumerable, and for each separate rotation there are only two fixed points,[28] the set Q is denumerable.

If $M \in K - Q$, and ρ_1 and ρ_2 are two distinct rotations in G, then $\rho_1(M) \neq \rho_2(M)$. In fact, in the contrary case we would have $\rho_2^{-1}[\rho_1(M)] = M$, and the point M would remain invariant under the rotation $\rho_1 \rho_2^{-1}$ which is known to be different from 1 due to our choice of the angle θ.

Having noted this, we denote the set of all points of the form $\rho(M)$, where ρ runs through the entire group G, by $H(M), M \in K - Q$.

If M and N are two distinct points of $K - Q$, then the sets $H(M)$ and $H(N)$ either are disjoint or they are identical. Therefore, the entire set $K - Q$ is *partitioned into classes* $H(M)$. From each of these classes, we select one representative and let X be the set of these representatives. Setting

$$R = \sum_{\rho \in A} \rho(X), S = \sum_{\rho \in B} \rho(X), T = \sum_{\rho \in C} \rho(X),$$

we obtain the sets which were mentioned in the formulation of the lemma.

In fact, we shall first establish that these sets do not intersect. But if the point P were in the intersection RS, then this would mean that

$$P = \rho_1(M) = \rho_2(N) \, [\rho_1 \in A, \rho_2 \in B, M \in X, N \in X].$$

The equality $M = N$ is impossible because the points M and N are not in Q. If however, $M \neq N$, then $\rho_1(M) = \rho_2(N)$ implies that $N \in H(M)$, and this is again impossible because the points M and N, being in X, are representatives of *distinct* classes H.

This means that $RS = 0$. Analogously, $RT = ST = 0$.

Further, $K - Q = R + S + T$. In fact, if $P \in K - Q$, then $P \in H(P)$. If the representative of the class $H(P)$, in X, is M, then $P \in H(M)$ and $P = \rho(M)$. According as ρ is in A, B or C, we have $P \in R$, $P \in S$ or $P \in T$, respectively.

Finally (and this is fundamental), we have that

$$\varphi(R) = S + T, \psi(R) = S, \psi^2(R) = T. \tag{6}$$

We shall verify, say, the first of these relations. If $P \in R$, then $P = \rho(M)$, where $M \in X$ and $\rho \in A$. But then

---

[27]This notation is slightly unsuitable. Namely, if, for instance, $\rho = \psi \varphi \psi^2$, then $\rho(M) = \psi^2\{\varphi[\psi(M)]\}$, because in the form $\rho = \psi \varphi \psi^2$, the rotation performed first is written on the left.

[28]By a known theorem in kinematics, every rotation about a point is a rotation about some axis passing through this point.

$$\varphi(P) = \varphi[\rho(M)] = (\rho\varphi)(M) \in S + T,$$

because, according to requirement I, we have $\rho\varphi \in B + C$. Conversely, if $P \in S + T$, then $P = \sigma(M)$, where $M \in X$, and $\sigma \in B + C$. We set $\rho = \sigma\varphi^{-1}$; then $\sigma = \rho\varphi$ and according to I, we have $\rho \in A$. Hence,

$$P = \varphi[\rho(M)],$$

where $\rho(M) \in R$. But this means that $P \in \varphi(R)$. Thus $\varphi(R) = S + T$. Analogously, (using only requirement II), the remaining relations (6) can be verified.

Thus, it remains only to verify that a partitioning of the group G into classes A, B, C, such that requirements I and II are satisfied, is possible.

To this end, we denote by G_n the set of those elements of G whose number of factors φ, ψ and ψ^2 is not greater than n when these elements are written in the reduced form. In particular,

$$G_1 = \{1, \varphi, \psi, \psi^2\}.$$

We put 1 into A, φ into B, ψ into B, and ψ^2 into C. This partitioning of the elements of G_1 satisfies both requirements I and II. In fact, if $\rho \in A$ (where ρ is in G_1), thi s means that $\rho = 1$, and then $\rho\varphi = \varphi \in B \subset B + C$. If, however, $\rho\varphi \in B + C$ (where both $\rho\varphi$ and ρ are in G_1), then we necessarily have $\rho\varphi = \varphi$ (because there are no other elements of the form $\rho\varphi$ in G_1, where $\rho \in G_1$, besides 1 and φ). Hence $\rho = 1 \in A$. Thus, requirement I is satisfied. We verify analogously that requirement II is satisfied. Namely, if ρ, $\rho\psi$ and $\rho\psi^2$ are all in G_1, then there can be only three possibilities for ρ:

$$\rho = 1, \rho = \psi, \rho = \psi^2. \tag{7}$$

Let one of the relations

$$\rho \in A, \rho\psi \in B, \rho\psi^2 \in C \tag{8}$$

be satisfied.

If this is the relation $\rho \in A$, then $\rho = 1$ and the second and third relations of (8) are fulfilled. If $\rho\psi \in B$, then it follows from possibilities (7) that $\rho = 1$ and this means that $\rho\psi^2 \in C$. Finally, if $\rho\psi^2 \in C$, then this also implies that $\rho = 1$.

Thus, the set G_1 is partitioned into classes A, B, C, both requirements I and II being satisfied. It is evident that if $\rho \in G_1$ and $\rho\varphi \bar{\in} G_1$, then to stipulate that requirement I be observed (in the set G_1) is devoid of meaning. The same is true with respect to the case when $\rho\varphi \in G_1$, but $\rho \bar{\in} G_1$ (e.g. $\rho = \psi\varphi$).

We shall assume that we have already partitioned the elements of the set G_n into classes A, B, C so that the requirements I and II are observed (again with the agreement that in the verification of the requirements, all[29] elements considered are in G_n).

We consider any element τ in $G_{n+1} - G_n$, so that τ consists of exactly $n + 1$ factors. There are three possibilities for τ:

$$\tau = \sigma\varphi, \tau = \sigma\psi, \tau = \sigma\psi^2, \tag{9}$$

where σ consists of exactly n factors, and, depending on which possibility of (9) holds, we have respectively

$$\sigma = \ldots\psi^m, \quad \sigma = \ldots\varphi, \quad \sigma = \ldots\varphi.$$

If $\sigma = \sigma\varphi (\sigma = \ldots\psi^m)$, then we put τ into the class B, A, A according as

$$\sigma \in A, \quad \sigma \in B, \quad \sigma \in C. \tag{10}$$

---

[29]For requirement I, these will be ρ and $\rho\varphi$; and for II, they will be ρ, $\rho\psi$ and $\rho\psi^2$.

If $\tau = \sigma\psi (\sigma = \ldots \varphi)$, we assume that

$$\tau \in B,\ \tau \in C,\ \tau \in A,$$

again depending on which relation of (10) holds.

Finally, if $\tau = \sigma\psi^2$ $(\sigma = \ldots \varphi)$, then, depending on which relation of (10) holds, we set

$$\tau \in C,\ \tau \in A,\ \tau \in B.$$

These statements uniquely define the partition of the entire group G into classes A, B, C. For the purpose of greater visualization, we give the scheme of partitioning G into classes A, B, C:

	Class τ		
Class σ	$\tau = \sigma\varphi$	$\tau = \sigma\psi$	$\tau = \sigma\psi^2$
A	B	B	C
B	A	C	A
C	A	A	B

(11)

As an example, we shall show how the elements of the set G_3 are partitioned into classes A, B, C:

	G_1	$G_2 - G_1$	$G_3 - G_2$	$G_4 - G_3$
A	1	$\psi\varphi,\ \psi^2\varphi,\ \varphi\psi^2$	$\varphi\psi\varphi$	. . . . .
B	$\varphi,\ \psi$		$\varphi\psi^2\varphi,\ \psi\varphi\psi,\ \psi^2\varphi\psi$	. . . . .
C	ψ^2	$\varphi\psi$	$\psi\varphi\psi^2,\ \psi^2\varphi\psi^2$	. . . . .

We shall show that in the partitioning of G into classes A, B, C, by the rule just indicated, requirements I and II (where, as already was stated, we shall assume that these requirements are already satisfied for G_n) will be satisfied for elements of the set G_{n+1}.

We take two elements ρ and $\rho\varphi$ belonging to the set G_{n+1}. We think of four cases:

1) $\rho \in G_n,\ \rho\varphi \in G_n$, 2) $\rho \in G_n, \rho\varphi \overline{\in} G_n$,

3) $\rho \overline{\in} G_n,\ \rho\varphi \in G_n$, 4) $\rho \overline{\in} G_n,\ \rho\varphi \overline{\in} G_n$.

In case 1), the two relations

$$\rho \in A, \rho\varphi \in B + C$$

are equivalent by assumption.

In case 2), ρ consists of exactly n factors where the last is ψ or ψ^2. If $\rho \in A$, then Table (11) shows that $\rho\varphi \in B \subset B + C$. Conversely, if $\rho\varphi \in B + C$, it is clear from Table (11) that $\rho\varphi \in B$ and $\rho \in A$.

Case 3) is possible if and only if

$$\rho = \sigma\varphi, \rho\varphi = \sigma, \sigma = \ldots\psi^m.$$

If $\rho \in A$, then it is clear from (11) that $\sigma \in B + C$, and conversely.

As to case 4), it is impossible. In fact, in this case ρ must consist of $(n + 1)$ factors where the last is not φ (because otherwise $\rho\varphi \in G_n$). Hence, $\rho = \ldots\psi^m$ and $\rho\varphi$ is not in G_{n+1}.

Thus, requirement I is satisfied.

Further, we take the elements $\rho, \rho\psi$ and $\rho\psi^2$ in G_{n+1}. Here, it is possible to distinguish three cases:

1) all three elements are in G_n;

2) none of the elements is in G_n;

3) some, but not all, of these elements are in G_n.

In case 1), requirement II is satisfied by assumption. Case 2) is impossible because in this case $\rho = \ldots\varphi$ and $\rho\psi \overline{\in} G_{n+1}$.

Consider case 3). We assume $\rho \in G_n$. Then we necessarily have

$$\rho = \ldots\varphi$$

(because otherwise $\rho\psi$ and $\rho\psi^2$ are also in G_n). If $\rho \in A$, then the first row of Table (11) shows that $\rho\psi \in B$, $\rho\psi^2 \in C$. Further, if (under the same assumption that $\rho \in G_n$) $\rho\psi \in B$, then $\rho \in A$, and if $\rho\psi^2 \in C$, then again $\rho \in A$. Thus, if $\rho \in G_n$, then requirement II is satisfied. Let $\rho \overline{\in} G_n$ but $\rho\psi \in G_n$. This means that $\rho = \sigma\psi^2$ (where $\sigma = \ldots\varphi$) and $\rho\psi = \sigma$. If $\rho \in A$, then $\sigma \in B$, and $\sigma\psi = \rho\psi^2 \in C$. In exactly the same way, if $\rho\psi \in B$, or $\rho\psi^2 \in C$, then $\rho \in A$. Thus, in this case also, requirement II is satisfied. Finally, let $\rho \overline{\in} G_n$ and $\rho\psi^2 \in G_n$. Then $\rho = \sigma\psi(\sigma = \ldots\varphi)$ and $\rho\psi^2 = \sigma$. If $\rho \in A$, then $\rho\psi^2 = \sigma \in C$ and $\rho\psi = \sigma\psi^2 \in B$ and each of the last relations in turn implies that $\rho \in A$. The proof is complete.

Remark. The lemma just proved is of a somewhat paradoxical nature. At the same time, its proof is not effective.[30] This brings to mind that we can avoid such paradoxes by applying only effective methods. For this reason, we must note that paradoxicality does not at all mean absurdity, but only some unusualness of the result. Therefore, no basis exists to assume that non-effective methods should be discredited because they led us to a paradox. Furthermore, and this is still more essential, we can construct such paradoxes as Hausdorff's lemma purely effectively. Here is a simple example[31] of this:

THEOREM. *There exists a plane set E which is representable in the form of the sum of two disjoint sets A and B, each of which is congruent to E.*[32]

To prove this, we consider two isometries in the plane of the complex variable z:

$$R(z) = e^i z, T(z) = z + 1.$$

Let E be a set consisting of the point $z = 0$ and also of all the points obtained from $z = 0$ with the aid of a finite number of isometries R and T. Each point of E is a polynomial with integral *non-negative* coefficients in the argument e^i, where, due to the transcendentality of the number e^i, the representation of a point of the set E in the form of such a polynomial is possible in only one way.

Having established this, we denote by A the set of those points of E which are representable as polynomials in e^i, *not having a free member*, and let $B = E - A$. Then $AB = 0, A + B = E$ and, moreover,

$$R(E) = A, T(E) = B.$$

---

[30]The *Zermelo axiom* (see Chapt. XIV) was used in the construction of the set X.

[31]Due to W. Sierpiński and S. Mazurkiewicz.

[32]The set E is not bounded and hence the unsolvability of the easy problem of measure for R_2 does not follow from the theorem (furthermore, E is denumerable).

VI. INDEFINITE INTEGRALS AND ABSOLUTELY CONTINUOUS SET FUNCTIONS

We shall now generalize several results of Chapt. XIII to functions with unbounded domains of definition. For the sake of simplicity, we shall consider only functions of one variable, defined[33] on $(-\infty, +\infty)$.

THEOREM 1. *A necessary and sufficient condition that a measurable function $f(x)$ be summable on every set of finite measure is that $f(x)$ be representable in the form*

$$f(x) = g(x) + h(x), \tag{1}$$

where $g(x)$ is summable on $(-\infty, +\infty)$, and $h(x)$ is measurable and bounded.

Proof. The sufficiency of the condition of the theorem is trivial. We shall prove its necessity. Thus, let $f(x)$ be summable on every set of positive measure. We shall first assume that $f(x) \geqslant 0$ and introduce the sets

$$E_k = E(f > 2^k) \qquad (k = 1, 2, 3, \ldots).$$

We shall show that at least one of these sets has finite measure. In fact, if we had $mE_k = +\infty$ for all k, then we would introduce measurable, bounded sets $e_1, e_2, e_3, \ldots$, such that

$$e_1 \subset E_1, me_1 = \frac{1}{2},$$

$$e_2 \subset E_2 - e_1, me_2 = \frac{1}{2^2},$$

. . . . . .

$$e_k \subset E_k - (e_1 + \ldots + e_{k-1}), me_k = \frac{1}{2^k},$$

. . . . . .

If $s = e_1 + e_2 + e_3 + \ldots$, then $ms = 1$, and hence $f \in L(s)$. At the same time, the sets e_k are pairwise disjoint, from which it follows that

$$\int_s f(x)\,dx = \sum_{k=1}^{\infty} \int_{e_k} f(x)\,dx > \sum_{k=1}^{\infty} 2^k me_k = +\infty,$$

and this contradicts the fact that $f \in L(s)$.

Thus, we can find an N such that

$$mE(f > 2^N) < +\infty.$$

Setting

$$g(x) = \begin{cases} f(x) & \text{for } f(x) > 2^N, \\ 0 & \text{for } f(x) \leqslant 2^N, \end{cases} \qquad h(x) = \begin{cases} 0 & \text{for } f(x) > 2^N, \\ f(x) & \text{for } f(x) \leqslant 2^N, \end{cases}$$

we obtain (1) with the required $g(x)$ and $h(x)$. If $f(x)$ can also assume non-negative values, then one must apply the part of the theorem already proved to $f_+(x)$ and $f_-(x)$.

---

[33] Everything asserted in the sequel of this appendix easily carries over to functions defined on a set A which is distinct from $(-\infty, +\infty)$. To this end, it is necessary only to extend the definition of these functions, setting them equal to zero in the exterior of A.

We denote by Γ the class of measurable functions which are summable on every set of finite measure. If $f \in \Gamma$, then to each set e, where $me < +\infty$, there corresponds a number

$$\Phi(e) = \int_e f(x)\,dx. \tag{2}$$

In other words, every function $f \in \Gamma$ generates a set function (2), defined on all sets e of finite measure. This function is countably additive[34] and absolutely continuous (in virtue of Theorem 1). The function $\Phi(e)$ is called the indefinite integral of the function $f(x)$.

THEOREM 2. *Every absolutely continuous, countably additive function $\Phi(e)$, defined on sets of finite measure, is the indefinite integral of some function in Γ.*

Proof. Since the function $\Phi(e)$ is defined, in particular, for all measurable and *bounded* sets e, the results of Chapt. XIII are applicable to it. Therefore, the symmetric derivative

$$f(x) = D_x f(e) = \lim_{h \to 0} \frac{\Phi([x-h, x+h])}{2h}, \tag{3}$$

exists almost everywhere and is a function summable on every measurable and bounded set E, where[35]

$$\Phi(E) = \int_E f(x)\,dx \tag{4}$$

for every set of this sort.

We now take an arbitrary set E of finite measure. It is representable in the form

$$E = A + B, \text{ where } A = E(f \geqslant 0), B = E(f < 0).$$

Each of the sets A and B is measurable and has finite measure. The set A can be represented in the form of a sum

$$A = \sum_{k=1}^{\infty} A_k,$$

where the sets A_k are measurable and *bounded.* An equality (4) can be written for each of these A_k, from which it follows that

$$\Phi(A) = \sum_{k=1}^{\infty} \int_{A_k} f(x)\,dx.$$

This proves the inclusion $f \in L(A)$ and the equality

$$\Phi(A) = \int_A f(x)\,dx.$$

The equality, with A replaced by B, is proved analogously. But then $f \in L(E)$ and (4) is valid. This completes the proof of the theorem, inasmuch as the set E is arbitrary.

---

[34] In the sense that for $E = \sum e_k$, where $e_k e_i = 0 (k \neq i)$ and under the condition that $mE < +\infty$, we will have $\Phi(E) = \sum \Phi(e_n)$.

[35] We extend the definition of $f(x)$, setting $f(x) = 0$ at those points where the limit (3) does not exist (these points form a set of measure 0).

COROLLARY. *The class of absolutely continuous, countably additive set functions, defined on sets of finite measure, coincides with the class of indefinite integrals of functions in* Γ.

In connection with this corollary, we may speak in Theorem 3, §2, Chapt. XIII, about functions in Γ rather than about summable functions.

THEOREM 3. *A necessary and sufficient condition that an absolutely continuous, countably additive function* $\Phi(e)$, *defined on sets of finite measure, be bounded is that it be the indefinite integral of a function summable on the entire real line.*

Proof. According to the preceding theorem, (2) is valid for $f \in \Gamma$. If $f \in L[(-\infty, +\infty)]$, then for an arbitrary set e of finite measure we have

$$|\Phi(e)| \leqslant \int_{-\infty}^{+\infty} |f(x)|\, dx,$$

which proves the sufficiency of the condition of the theorem. Conversely, if for every such set e we have $|\Phi(e)| \leqslant K$, then a similar inequality can be written for the sets $e(f \geqslant 0)$ and $e(f < 0)$, from which it follows that

$$\int_e |f(x)|\, dx \leqslant 2K. \tag{5}$$

In particular, if we set $e = [-n, +n]$ and pass to the limit as $n \to +\infty$ in (5), we obtain

$$\int_{-\infty}^{+\infty} |f(x)|\, dx \leqslant 2K,$$

which completes the proof of the theorem.

VII. THE ROLE OF RUSSIAN AND SOVIET MATHEMATICIANS IN THE DEVELOPMENT OF THE THEORY OF THE FUNCTIONS OF A REAL VARIABLE

As was indicated already in Chapter I, the theory of functions of a real variable is one of the youngest branches of mathematics. Its history starts with the beginning of the twentieth century when the methods of set theory were placed at the foundation of the study of functions. In this connection, two great tendencies of development appeared rather early: the *descriptive* theory of functions which does not use the concept of *measure* of a set, and the *metric* theory of functions where, in contrast, the concept of measure is basic. Somewhat aside from these directions, yet a third great tendency is found—it is the *constructive* theory of functions, which is concerned with problems of the approximate representation of functions. In the preceding pages of this book, the constructive theory of functions was scarcely touched upon; for that reason we shall not stop to consider it in the present survey. We shall only note that it has its source in the remarkable works of the great mathematician P. L. Chebyshev and that the entire modern development of the constructive theory of functions was determined by the profound researches of S. N. Bernstein. In the sequel of this appendix, when we are

speaking of the theory of functions of a real variable, we shall not include the constructive theory of functions.

The beginning of the systematic development of the theory of functions of a real variable in Russia is connected with the name of D. F. Egorov, who proved his renowned theorem on convergent sequences of measurable functions in 1911. Immediately after this, in 1912, the work of N. N. Luzin appeared, in which his theorem on the construction of an arbitrary measurable function was proved; this theorem is considered basic in the metric theory of functions, and justifiably so. In 1915, the doctoral dissertation of N. N. Luzin: *Integrals and Trigonometric Series* was published. This work, full of new and fresh ideas, served as point of departure for an extensive series of works by Moscow mathematicians. Beginning with the twenties of this century, the Moscow school of function theory attained the first place in the world. Besides works of the Moscow school, we can note activity in other mathematical centers in the USSR. In particular, a significant group of students is working in Leningrad under the direction of G. M. Fichtenholz; in Odessa, under the guidance of M. G. Krein, functional analysis and its applications are being very intensively developed.

After these introductory remarks, we shall now proceed to a very brief and incomplete[36] survey of Soviet works on the theory of functions of a real variable.

* * *

We shall first discuss works in the descriptive theory of functions.

Every non-denumerable closed set has a perfect subset and therefore its power equals the power of the continuum. The question naturally arises, what is the power of the more complicated Borel sets? This problem was solved in 1916 by P. S. Aleksandrov who proved that *every non-denumerable Borel set contains a perfect subset.* To prove this fact, P. S. Aleksandrov established a special method of defining the Borel sets. Namely, let us assume that to every finite collection of natural numbers $(n_1, n_2, \ldots, n_k)$ there corresponds a closed set $F_{n_1 n_2 \ldots n_k}$. Then we can form the set

$$S = \sum F_{n_1} F_{n_1 n_2} F_{n_1 n_2 n_3} \ldots, \qquad (*)$$

where the summation extends over all infinite sequences $(n_1, n_2, n_3, \ldots)$ of natural numbers. By varying the method of corresponding the sets $F_{n_1 n_2 \ldots n_k}$ to the collections $(n_1, n_2, \ldots, n_k)$, we shall obtain different sets S. It turns out that every Borel set is representable in the form (*). P. S. Aleksandrov used this fact to prove his theorem. In 1917, M. Ya. Suslin (1894-1919) proved that *the class of sets which can be given in the form* (*) *is essentially more extensive than the class of Borel sets.* Suslin called the sets of this new class *analytic* sets or A-sets. Suslin himself only began the investigation of these sets. The further development of the theory of these sets is connected, chiefly, with the name N. N. Luzin, who included them in the theory of the more general, so-called *projective*, sets. The works of P. S. Novikov, A. A. Lyapunov, and others, are related to this. We introduce some results from this area.

1. *A necessary and sufficient condition that an analytic set be a Borel set is that its complement also be an analytic set* (M. Ya. Suslin).

2. *A necessary and sufficient condition that an analytic set be a Borel set is that it*

[36]The reader will find an unusually complete discussion of the problem in the monumental work *Mathematics in the USSR during the Thirty Years* 1917-1947, Gostekhizdat, 1948.

be representable in the form (*) *where the terms are pairwise disjoint* (N. N. Luzin).

3. *The class of analytic sets coincides with the class of projections on a line of plane Borel sets* (N. N. Luzin).

4. *All analytic sets are measurable* (N. N. Luzin).

5. *Every non-denumerable analytic set contains a perfect subset* (P. S. Aleksandrov).

6. *Two disjoint analytic sets* A_1 *and* A_2 *are separated by Borel sets, i.e. there exist two Borel sets* B_1 *and* B_2 *such that* $B_1 \supset A_1, B_2 \supset A_2$, *and* $B_1 B_2 = 0$ (N. N. Luzin).

7. *Two disjoint sets which are the complements of analytic sets might not be separated by Borel sets* (P. S. Novikov).

The theory of analytic sets served as a starting point for the creation of a series of more general theories. Above all, in connection with the representation of analytic sets in the form (*), there was given a more general definition of A-operations. Namely, let T be a class of sets. Assign to each finite collection $(n_1, n_2, \ldots, n_k)$ of natural numbers a definite set $E_{n_1 n_2 \ldots n_k} \in T$ and set

$$S = \sum E_{n_1} E_{n_1 n_2} E_{n_1 n_2 n_3} \ldots,$$

where the summation extends over all infinite sequences of natural numbers. We say that the set S is the result of A-operations on the sets of the class T. In particular, analytic sets are the result of A-operations on closed sets. It turns out that carrying out A-operations on analytic sets does not take us out of the class of these sets. However, the result of A-operations on sets which are the complements of analytic sets is, in general, a set of an essentially new sort. By alternating the A-operation with the formation of complements, P. S. Aleksandrov formed an entire sequence of more and more comprehensive classes of sets. These sets (i.e. the so-called C-sets) were the object of study by P. S. Aleksandrov himself, who proved their topological invariance, and by E. A. Selivanovsky.

On the other hand, N. N. Luzin also constructed a transfinite classification of sets, by alternating the operation of projection with the operation of forming complements. The theory of sets, appearing in this classification, of the so-called "projective sets," turned out to be very difficult. In particular, the question of the measurability of these sets has not yet been explained.

Starting from the concept of an A-operation, A. N. Kolmogorov laid the foundation for a general theory of operations, which bring definite new sets into correspondence with the given sets. The theory of such operations (they are called analytic operations) made essential headway in the works of the Leningrad mathematicians L. V. Kantorovič and E. M. Livenson. In particular, they proved that all C-sets are already contained in the second projective class of N. N. Luzin.

Of other works on the descriptive theory of functions, we mention the investigations of L. V. Keldysh. Her works include a series of results related to the variegated properties of Borel sets and Baire functions. In particular, she was the first to construct a purely arithmetical example of a function of the fourth Baire class. Later, L. V. Keldysh constructed such examples for all finite Baire classes. Space does not allow us to stop to consider other Soviet works on the descriptive theory of functions.

* * *

We shall next consider works related to the theory of measurable functions and to differentiation and integration.

Soviet works on the metric theory of functions are so numerous and variegated

that even a most cursory survey requires their subdivision into several leading directions. It is natural to begin with the works devoted to the general theory of measurable functions and their differentiation and integration. Here, above all, we ought to stop to consider the monumental work of A. Ya. Khinchin, in which a series of important properties of measurable functions are established. Already in §5, Chapter IX we saw that every measurable function is continouous at almost every point if we allow ourselves to disregard sets having zero density at this point.

In this relation, A. Ya. Khinchin suggested that one say that a property holds at the point x *asymptotically*[37] if it remains after removing a set having zero density at the point x. The concept of asymptotic direction of a function at a point, asymptotic differentiability, and so on, are introduced in this way.

A. Ya. Khinchin established the following interesting analogue of the Luzin Theorem: *a function, directed asymptotically almost everywhere, can be transformed into a continuous function with a finite number of extrema by modifying it on a set of arbitrarily small measure.* It follows directly from this that *such a function is asymptotically differentiable almost everywhere.* It would be natural to attempt to prove that *every* measurable function is asymptotically differentiable. But this is not so, as A. Ya. Khinchin showed. For an arbitrary measurable function, we can only guarantee that it is almost everywhere either asymptotically differentiable or its upper asymptotic derivatives from the right and left equal $+\infty$ and that its lower asymptotic derivatives equal $-\infty$.

The works of V. V. Stepanov are intimately related to the investigations of A. Ya. Khinchin. V. V. Stepanov established that *a function which is almost everywhere asymptotically continuous is necessarily measurable,* and for functions of several variables he proved the remarkable and unexpected theorem which states that *the existence of only one asymptotic partial derivative on any measurable set implies the existence of the asymptotic total derivative almost everywhere on this set.*

The works of G. P. Tolstov, Yu. B. Germeier, M. A. Zarecki, and others, are related to the aforementioned works.

A series of works was devoted to the special problems of the differentiation of arbitrary functions. We mention the works of V. N. Veniaminov, L. V. Kantorovič, I. M. Liberman, I. G. Petrovsky, and others.

Going over to questions connected with the *problem of integration,* we note that already in 1916, A. Ya. Khinchin gave a new definition of integral which was more general than the definition of Denjoy and which enables one to reconstruct the primitive function simply by knowing its asymptotic derivative. In the great work of A. N. Kolmogorov, the very concept of integral was subjected to profound analysis and a very general definition of integral was given which contains almost all definitions known earlier as special cases. V. I. Glivenko gave a still more general concept of integral. Of other works on the theory of the integral, we mention a work of P. S. Aleksandrov in which the equivalence of the two definitions of the integral, suggested independently by Denjoy and Perron, is proved and also the work of P. I. Romanovsky, where a very natural construction of the theory of the Denjoy integral was suggested.

* * *

[37] In accordance with the terminology adopted in this book, it would be natural to say that the property considered holds *approximately.*

A series of results of fundamental significance was obtained by Soviet mathematicians in the theory of trigonometric series. Here, chronologically, the first were the results of N. N. Luzin, discussed in his dissertation which was already mentioned above. One of these results was introduced in §4, Chapter X. Of the others, we introduce the following theorem: *Every measurable and almost everywhere finite function* $f(x)$, *defined on* $[-\pi, \pi]$, *can be represented as a trigonometric series which is Abel-Poisson summable and Riemann summable to it everywhere.* But then N. N. Luzin posed the question whether it is possible to replace *summability* of the series in this formulation by *convergence.* This question did not lend itself to solution for 25 years. In 1940, D. E. Menshov solved the problem by proving that *for every measurable and almost everywhere finite function, defined on* $[-\pi, \pi]$, *there exists a trigonometric series converging to it almost everywhere.*

In connection with these results, the question arises *how many* trigonometric series representing a given function exist? It was already proved by Cantor that in the case of a trigonometric series, which converges everywhere, a second trigonometric series cannot exist which also converges everywhere to the same sum. If we form the difference of the trigonometric series considered, the Cantor theorem can be formulated as follows: if a trigonometric series converges everywhere and its sum is zero, then all its coefficients are zero. Cantor established the more general proposition: if a trigonometric series converges to a sum, equal to zero, everywhere except at most on a finite set E of "points where the function is not known," then all its coefficients are zero (so that it also converges to zero on the set E). These facts led to the formulation of the *general problem of uniqueness* in the theory of trigonometric series. This problem is formulated in the following manner: a point set E, lying on $[-\pi, \pi]$, is said to be a set of type M, if there exists a trigonometric series not all of whose coefficients equal zero and which converges to the sum, equal to zero, everywhere in the exterior of the set E. But if such a series does not exist, then the set E is said to be a set of type U. In other words, E is a set of type U if every trigonometric series, which converges to zero everywhere in the exterior of E, has all coefficients equal to zero. It is clear that a subset of a set of type U is itself a set of the same type. Roughly speaking, the "more opulent" the set the more likely is it to be a set of type M, and the "more meager" the set the more likely will it be of type U. The example of Cantor, which was introduced above, signifies that all finite sets are of type U. It is not difficult to see[38] that every set of positive inner measure is a set of type M.

In connection with the above discussion, and also in view of Young's theorem, which states that all denumerable sets are sets of type U, the conviction was dominant that *all* sets of measure zero are sets of type U. So the example of a set of measure zero and type M, published by D. E. Menshov in 1916, was all the more unexpected. After this, the opinion arose that the question whether a given set is an M or a U set is connected

---

[38] In fact, let $E \subset [-\pi, \pi]$ and $m_* E > 0$. Then there exists a closed set F of positive measure contained in E such that the points π and $-\pi$ are both in F or neither is in F. The Fourier series of the characteristic function $f(x)$ of the set F will converge to zero everywhere in the exterior of F and, hence, a fortiori, in the exterior of E. At the same time, its free member is

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx = \frac{mF}{2\pi} > 0.$$

with the *power* of the set. In 1921, N. K. Bari showed this was not so; she found an entire class of perfect U-sets. The question of which sets are U-sets and which are M-sets is not yet solved although a number of results have been obtained in this direction. We mention, for example, N. K. Bari's theorem that *the sum of a denumerable family of closed U-sets is again a U-set.*

Questions of uniqueness were taken up by N. K. Bari, D. E. Menshov, and by many other Soviet mathematicians. We mention the works of A. A. Shneider, V. Ya. Kozlov, Ya. V. Bykov, E. A. Nersesov, and others.

Of other works on the theory of trigonometric series, the example by A. N. Kolmogorov of a *summable function with an everywhere divergent Fourier series* is of fundamental significance. The question of the existence of a continuous function and also of functions in L_2 with Fourier series which diverge on a set of positive measure still remains open.

The theorem discovered in 1926 by A. N. Kolmogorov and G. A. Seliverstov, and independently by A. I. Plessner, that *the convergence of the series*

$$\sum_{n=1}^{\infty}(a_n^2+b_n^2)\ln n$$

implies the convergence almost everywhere of the series

$$\sum_{n=1}^{\infty}(a_n\cos nx+b_n\sin nx)$$

is of significance.

We shall not stop to consider other works of Soviet mathematicians which deal with trigonometric series. We only note that in these works all phases, without exception, of the theory of such series were advanced: the questions of the absolute convergence (N. N. Luzin, S. N. Bernstein, V. V. Nemytsky, I. M. Gelfand, V. Ya. Kozlov, and others), questions of summing trigonometric series (S. N. Bernstein, S. M. Lozinsky, I. P. Natanson, S. M. Nikolsky, F. I. Kharshiladze, and others), theory of conjugate series (A. N. Kolmogorov, A. I. Plessner, I. I. Privalov, V. I. Smirnov), questions of the characteristics of the Fourier series of functions of a given class (G. M. Fichtenholz, F. I. Kharshiladze, I. P. Natanson), and many others.

* * *

The theory of orthogonal functions occupies an extensive place in Soviet investigations. Here, the most significant are the distinguished works of D. E. Menshov.

Plancherel already established that for an arbitrary orthogonal system $\{\omega_n(x)\}$, the convergence of the series

$$\sum_{n=1}^{\infty}c_n^2\log^3 n$$

implies the convergence almost everywhere of the series

$$\sum_{n=1}^{\infty}c_n\omega_n(x).$$

D. E. Menshov and H. Rademacher, independently of one another, showed that the factor $\log^3 n$ can be reduced to $\log^2 n$, and D. E. Menshov established that further reduction is not possible, i.e. *for an arbitrary function* $w(n)$, *for which*

$$\lim_{n\to\infty} \frac{w(n)}{\log^2 n} = 0,$$

we can construct an orthogonal system $\{\omega_n(x)\}$ *and numbers* c_n *such that the series* $\sum w(n)c_n^2$ *converges and the series* $\sum c_n\omega_n(x)$ *diverges on a set of positive measure.*

Comparing this result with the aforementioned theorem of A. N. Kolmogorov, G. A. Seliverstov and A. I. Plessner leads to the question of whether for every *uniformly bounded* orthonormal system, the convergence almost everywhere of the series $\sum c_n\omega_n(x)$ follows from the convergence of the series $\sum c_n^2 \log n$? D. E. Menshov gave this question a negative answer.

The work of N. Ya. Vilenkin, and the work done by A. N. Kolmogorov together with A. Ya. Khinchin, are related to these investigations. In the first of these papers, it is shown that *the convergence of the series* $\sum c_n^2 \log n$ *implies the convergence almost everywhere of the series* $\sum c_n\omega_n(x)$, *where* $\omega_n(x)$ *are the orthonormal Walsh functions.* In the second paper, *the divergence almost everywhere of an orthonormal series in Rademacher functions* $\sum c_n r_n(x)$ *under the condition that the series* $\sum c_n^2$ *diverges is proved.*

Besides the questions of the convergence of orthogonal series, D. E. Menshov investigated questions of summability of these series by Cesàro processes and general Toeplitz processes, and also with the study of the effect of the rearrangement of orthogonal functions on the convergence and summability of the series of these functions.

I. I. Privalov investigated the absolute convergence of orthogonal series. He proved that *the absolute convergence almost everywhere of the orthogonal series* $\sum c_n\omega_n(x)$ *with respect to uniformly bounded functions implies the convergence of the series* $\sum |c_n|$. But if we omit the condition of uniform boundedness of the functions of the system, it is impossible even to guarantee that the coefficients c_n tend to zero.

In Kazan, B. M. Gagayev studied various questions of the theory of orthogonal series. In particular, he solved N. N. Luzin's problem of what sort of orthonormal system is invariant with respect to differentiation. His results on this topic were extended by B. V. Gnedenko. In later works, B. M. Gagayev explained what functions which are orthogonal with respect to one weight, transform, under differentiation, into functions which are orthogonal with respect to another weight.

A number of authors were specially studying the questions of *completeness* of orthonormal systems. At the beginning of the century, V. A. Steklov investigated these questions to a great extent and here very essential results are due to him. In particular, he established the completeness of many concrete orthogonal systems, for example, those formed by Laguerre polynomials and by Hermite polynomials. In Chapter VII we saw that in L_2 the properties of completeness and closure of orthogonal systems are equivalent. In 1926, G. M. Fichtenholz established that *in classes more restricted than* L_2, *for example in the class of functions integrable in the Riemann sense, these properties are no longer equivalent* (closure always implies completeness, but not conversely!).

N. K. Bari considered the question of what properties of orthogonal systems are "stable" in some sense, i.e. are retained in the transition from an orthogonal system possessing this sort of property to a "proximate" system. In this connection, the "proximity" of two orthogonal systems can be understood in various ways. Let us introduce, for example, the following result of N. K. Bari: *If* $\{\omega_n(x)\}$ *and* $\{\varphi_n(x)\}$ *are two orthonormal systems, proximate in the sense that*

$$\sum_{n=1}^{\infty} \int_a^b [\omega_n(x) - \varphi_n(x)]^2 dx < +\infty,$$

then these systems are simultaneously complete or not complete. Later, N. K. Bari extended some of her results on "stability properties" from the case of orthonormal systems also to other systems of functions. The works of K. I. Babenko, M. M. Grinblyum, V. Ya. Kozlov, and also the work of M. G. Krein, D. P. Milman and M. A. Rutman (done together), are related to these works of N. K. Bari.

N. P. Romanov used the theory of orthonormal systems to obtain an entire series of interesting arithmetic identities.

As in other branches, we are far from exhausting the variegated Soviet works on the theory of orthogonal series, but space does not permit us to go into further detail.

* * *

We shall now discuss other works on the metric theory of functions.

In 1922, G. M. Fichtenholz[39] established that *the superposition of two absolutely continuous functions will be an absolutely continuous function if and only if the total variation of this superposition is finite.* The result of G. M. Fichtenholz turned out to be the starting point for a series of interesting investigations by Soviet and Polish mathematicians. A little on this topic was already discussed above (cf. S. S. Banach-M. A. Zarecki's Theorem). Of other results related to the problem touched upon, we introduce the theorem of N. K. Bari and D. E. Menshov: *A necessary and sufficient condition that a continuous function be representable in the form of a superposition of two absolutely continuous functions is that this function map all sets of sufficiently small measure into sets of arbitrarily small measure.* It follows from this, among other things, that *the formation of the superposition of three absolutely continuous functions does not yield new functions.* Later, N. K. Bari discovered the remarkable fact that *every continuous function is representable in the form of a sum of three superpositions of absolutely continuous functions, but there exist continuous functions which are not representable in the form of the sum of two such superpositions.* A number of results was also obtained by N. K. Bari in relation to the superposition of functions of finite variation.

Much attention was given by Soviet mathematicians to the *theory of integrals which depend on a parameter.* Already in 1918, the dissertion of G. M. Fichtenholz was devoted to these problems; a number of interesting results are contained in this work. Thus, for example, an exceptionally simple condition for the validity of the interchange of two Riemann integrations was given there (this condition was found by G. M. Fichtenholz in 1910). Later, G. M. Fichtenholz returned to these questions many times.

Singular integrals are closely related to integrals depending on a parameter. In the thirties, a number of Soviet works were devoted to this topic. The results of P. I. Romanovsky and D. K. Faddeyev were discussed in §2, Chapter X. We introduce further the following results of I. P. Natanson:

1. *A necessary and sufficient condition that for an arbitrary measurable function $f(t)$, approximately continuous at the point x, we have*

[39]See §3, Chapt. IX.

$$\lim_{n\to\infty} \int_a^b f(t)\,\Phi_n(t,x)\,dt = f(x) \qquad (*)$$

is that for every set E of density 1 *at the point x,*

$$\lim_{n\to\infty} \int_E \Phi_n(t,x)\,dt = 1.$$

2. *Kernels* $\Phi_n(t,x)$ *do not exist such that the equality* (*) *holds for every summable function which is approximately continuous at the point x.*

I. P. Natanson also investigated questions of non-local convergence of singular integrals. Of other works on singular integrals, we mention the works of S. M. Lozinsky and F. I. Kharshiladze.

In the forties, a series of new works by B. I. Korenblyum, S. G. Krein and B. Ya. Levin were devoted to the theory of singular integrals, where the last two authors succeeded in giving the question a fresh impulse due to the methods borrowed from functional analysis.

Finally, we mention numerous works on functions of several variables. Here, besides the theorem of V. V. Stepanov, already mentioned above, very interesting and important results were obtained by G. P. Tolstov, A. S. Kronrod, and A. D. Myshkis. Here, the interesting works of I. Ya. Verchenko and A. N. Kolmogorov on the nature of the discontinuities of functions of two variables are relevant.

* * *

In conclusion of this brief survey, we shall stop very briefly on functional analysis. This important branch of mathematics was scarcely touched upon in our book and, therefore, we shall restrict ourselves here to only the most general facts.

At the present time, we can ascertain the presence in the USSR of three large groups of researchers in functionals analysis: the Moscow, Leningrad, and Odessa groups.

The first Moscow works on functional analysis were related to its applications in the domain of classical analysis (the works of V. V. Nemytsky on integral equations, A. N. Tikhonov on differential equations, L. A. Lyusternik and L. G. Shnirelman on the calculus of variations). Later, there occurs a transition to the study of the theory of linear operators in Hilbert space (A. I. Plessner, I. M. Gelfand), and still later, in Moscow, I. M. Gelfand and his co-researchers D. A. Raikov, G. E. Shilov, and others, formed a large new chapter of functional analysis—the "theory of normed rings" or "Banach algebras"—(here we must mention the important work of V. A. Ditkin, which appeared before the creation of the general theory of normed rings). A number of isolated Moscow works served as the initial point of numerous further investigations. Such are, for example, the renowned works of A. N. Kolmogorov on the normability of topological spaces and on compactness in L_p.

In Leningrad, work on functional analysis began with the simultaneous investigations of G. M. Fichtenholz and L. V. Kantorovič on the general form of a linear functional in the space of bounded functions. Later, G. M. Fichtenholz studied abstract spaces from the viewpoint of the convergence defined in them, and L. V. Kantorovič published an extensive series of works in which the theory of abstract spaces of a special form, having extensive applications, was created; these were the so-called "semi-ordered

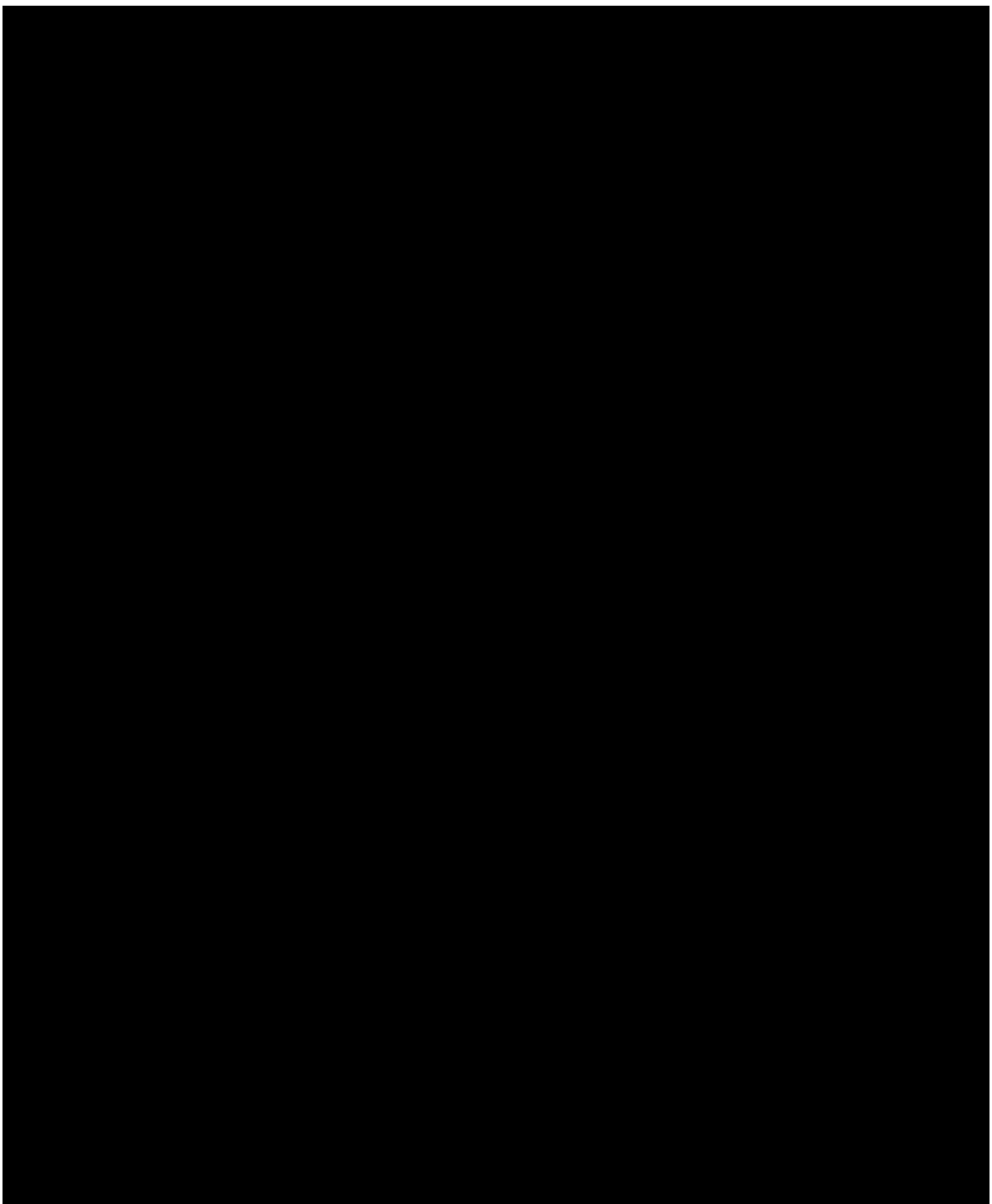
spaces." In the subsequent development of the theory of these spaces, the students of L. V. Kantorovič: B. Z. Vulikh, A. G. Pinsker, and A. I. Yudin (who died at the front in the Second World War), participated to a great extent. In recent years, L. V. Kantorovič was concerned with the systematic application of functional analysis to the theory of approximate calculations. Here, he obtained results of extraordinary generality and power.

Of other Leningrad works on functional analysis, we mention the investigations of M. K. Gavurin, which are devoted to non-linear operators and their analytic representation, the works of A. D. Aleksandrov and A. A. Markov on functionals of functions defined in topological spaces, and the interesting article by G. P. Akilov relating to the problem of the extension of linear operators.

M. G. Krein is the head of the Odessa school of functional analysis. He arrived at the general problems of functional analysis, starting from his investigations on the theory of matrices and the theory of integral equations, and also proceeding from his works on the constructive theory of functions, which were done in collaboration with N. I. Akhiezer. The students and co-researchers of M. G. Krein are V. R. Gantmakher, D. P. Milman, M. A. Rutman, V. L. Shmulyan, and others. At the present time, the questions of the geometry of abstract spaces, mainly spaces of Banach type, are being investigated intensively at this school. The results obtained find numerous and various applications. Thus, for example, the concept of a *strictly normed* Banach space, introduced by M. G. Krein, turned out to be very useful in problems dealing with the theory of approximations. The works of S. G. Krein and B. Ya. Levin on singular integrals, mentioned above, also carry the stamp of the ideas of the Odessa school of functional analysis.

In speaking of functional analysis, it is impossible not to mention also the work accomplished in Kiev by N. M. Krylov and N. N. Bogolyubov in connection with the qualitative theory of differential equations.

In bringing this survey to a close, I ought again to emphasize its incompleteness. Not only has mention of many isolated works been omitted from it, but also some significant directions of investigation are not even touched upon. Thus, for example, we did not touch upon the theory of almost periodic functions (works of V. V. Stepanov, B. M. Levitan, A. S. Kavanko, and others), questions bordering on the theory of functions of a complex variable (works of N. N. Luzin, I. I. Privalov, V. V. Golubev, A. I. Plessner, V. I. Smirnov, G. M. Fichtenholz, A. I. Markushevič, and others), questions pertaining to the foundations of set theory (I. I. Zhegalkin, D. A. Bochvar, P. S. Novikov, A. I. Maltsev), and many other works. Nonetheless, from what has been introduced, it is already clear that Soviet works on the theory of functions of a real variable not only are a valuable contribution to all branches of this science, but, to a significant degree, determine all its modern features.



INDEX

Abel condition, 30
Abel-Poisson summation method, 28
Abel's
  lemma, 29
  theorem, 30
Absolutely continuous set functions, 94, 246
Abstract definition of integral, 169
Addition of order types, 111
Additive set function, 94
Affine transformation, 72
Akhiezer, 257
Akilov, 257
Aleksandrov, 81, 145, 157, 185, 249, 250, 251
Aleksandrov-Looman theorem, 185, 189
Alephs, 122
  transfinite, 122
Analytic sets, 249
Antinomy, Burali-Forti, 117
Approximate continuity
  point of, 16
  derivative, 157
Arc
  Euler polygonal, 210
  length of a curve, 225
Arzelà-Ascoli theorem, 207
Asymptotic derivative, 157
Axiom, Zermelo's, 124

Babenko, 255
Baire
  classes, 128
  classification, 128
  functions, 43, 128
Baire's theorem, 143, 146
Banach, 195
  algebras, 256
  spaces, 195
Banach's
  fixed-point principle, 216
  theorem, 217
Bari, 54, 57, 253, 254, 255
Bari-Menshov theorem, 255
Bari's theorem, 253
Bernstein, 248, 253
  singular integral, 58
Birnbaum-Orlicz theorem, 232
Bochvar, 257
Bogolybov, 257
Bolzano-Weierstrass theorem, 60
Borel's theorem, 61
Bounded
  set, 199
  solution, 222
Burali-Forti antinomy, 117
Bykov, 253

Cantor, 54
Cantor-Baire stationary principle, 145
Cantor-Labesgue theorem, 34
Cantor's
  perfect set, 148
  theorem, 51, 52, 55
Center of a circle, 60
Cesàro summation method, 22
Cesàro-Fejér summation method, 28
Change of variable, 234
Characteristic function, 152
Chebyshev, 248
Closed
  circle, 60
  sets, 60, 199
  sphere, 199

Compact
set, 199
space, 199
Compactness, 198, 202
Comparison
of ordinal numbers, 115
of Perron and Lebesgue integrals, 165
Complementary in the Young sense, 232
Complete
metric space, 194
trigonometric system, 25
Completely
additive set function, 94
regular system, 222
Conditions for compactness, 202
Constructive theory of functions, 248
Continuum hypothesis, 127
Contractible family, 103
Contraction operator, 217
Convergence, weak, 15
Convergent in itself, 193
Convex
downward, 38
functions, 11, 36, 230
majorant, 20
Countably additive set function, 94
Covering
in the Vitali sense, 70
of a set, 61
C-set, 250

De la Vallée Poussin, 58, 234
singular integral, 58
Poussin's theorem, 40
Denjoy, 54, 157
Denjoy integral
in the restricted sense, 175
in the wide sense, 189
Denjoy-Khinchin integral, 189
Denjoy-Perron integral, 175
Derivative
approximate, 157
asymptotic, 157
lower, 158
Schwarz, 36
second Schwarz, 37
symmetric, 95
upper, 158
with respect to a family of measurable sets, 103
Derived number
second Schwarz, 38
set, 60
Descriptive theory of functions, 248
Determinant of a transformation, 72
Diameter of a set, 103, 201
Differentiation of the indefinite integral, 100
d-point, 57
Dirichlet singular integral, 22
Dirichlet's function, 148, 149
Distance, 60
in a metric space, 193
Ditkin, 256
Du Bois-Reymond-de la Vallée Poussin theorem, 52

Egorov, 249
ε-net, 200
Equicontinuous functions, 206
Euler
function, 210
polygonal arc, 210

Faddeyev, 21, 255
Faddeyev's theorem, 20
Family of sets contractible to a point, 103
Fatou, 58
Fejér singular integral, 23
Fejér-Lebesgue theorem, 25
Fichtenholz, 22, 28, 92, 249, 254, 255, 256, 257
First
category set, 142
class of functions, 128, 139
number class, 119
Schwarz derivative, 36
Fixed point of an operator, 217

Fourier series, 21, 29
Fubini's theorem, 86
Function
  absolutely continuous set, 94, 246
  Additive set, 94
  Baire, 43, 128
  characteristic, 152
  convex, 11, 36, 230
    downward, 38
  Euler, 210
  induced, 143
  integrable in the Perron sense, 159
  lower semi-continuous, 150, 152
  measurable, 80
  (P)-integrable, 159
  representation of a (by a singular integral), 15
  Riemann, 50
  semi-continuous, 149
  set, 94
  Steklov, 214
  subgraph of a, 83
  truncated, 87
  universal, 139
  upper semi-continuous, 150, 152
Functional analysis, 193
Functions
  of the first class, 128, 139
  of the second class, 128

Gagayev, 133, 254
Gantmakher, 257
Gavurin, 257
Gelfand, 253, 256
General principle of choice, 124
Generalizations of the Lebesgue integral, 157
Geometric interpretation of the Lebesgue integral, 83
Germeier, 251
Glivenko, 251
Gnedenko, 254
Golubev, 257
Grinblyum, 255

Hake, 157, 178
Hake's theorem, 178, 185
Hausdorff's theorem, 238
Hölder's inequality, 234

Indefinite
  integral, 100, 246
  Perron integral, 163
Induced function, 143
Inequality
  Hölder's, 234
  Jensen integral, 46
    summation, 44
  triangle, 193
  Young's, 231
Integral,
  abstract definition of, 169
  Bernstein, 58
  de la Vallée Poussin, 58
  Denjoy (in the restricted sense), 157, 175
    (in the wide sense), 157, 189
  Denjoy-Khinchin, 157, 189
  Denjoy-Perron, 157, 175
  Dirichlet, 22
  Fejér, 23
  indefinite, 100, 246
    Perron, 163
  Kantorovič, 58
  Landau, 58
  Lebesgue, 83
    (generalizations of the), 157
  Perron, 158
  Poisson, 29, 58
  singular, 12
  Weierstrass, 20
Interchanging order of integration, 90
Interior point, 62, 199
Invariants under isometries, 70
Isolated point, 60
Isometry, 70

Jensen integral inequality, 46
  summation inequality, 44
Jordan's theorem, 226

Kaczmarz, 27
Kantorovič, 22, 28, 58, 139, 250, 251, 256, 257
  singular integral, 58
Kavanko, 257
Keldysh, 250
Kernel, 12
Kharshiladze, 253, 256
Khinchin, 157, 251, 254
Koenig, 127
Kolmogorov, 27, 212, 228, 250, 251, 253, 254, 256
Kolmogorov's theorem, 26, 27, 215
Korenblyum, 256
Kozlov, 159, 253, 255
Krein, 249, 255, 256, 257
Kronrod, 256
Krylov, 257

Lacunary
  sequence, 26
  trigonometric series, 27
Landau singular integral, 58
Lebesgue, 58, 124, 133
  integral, 83
    change of variable in the, 234
    generalizations of, 157
  point, 16
Lebesgue's theorem, 13, 15, 137, 141
Legitimate system of classes, 169
Lemma
  Abel's, 29
  Natanson's, 16
  Riemann's, 48
  Schwarz's, 55
Levin, 256, 257
Levi's theorem, 88
Levitan, 257
Liberman, 251
Limit
  element, 198
  inferior, 134, 149
  of a sequence, 193
  point, 60, 198
  superior, 134, 149
Linear
  space, 194
  spaces which are not normed, 197
Livenson, 250
Looman, 157, 185
Lower
  derivative, 158
  semi-continuous function, 150, 152
Lozinsky, 253, 256
Luzin, 83, 249, 250, 252, 253, 254, 257
Luzin-Denjoy theorem, 35
Lyapunov, 249
Lyusternik, 256

Majorant, convex, 20
Maltsev, 257
Markov, 119, 257
Markushevič, 257
Mazurkiewicz, 245
Measurability as invariant under isometries, 70
Measurable
  function, 80
  functions of several variables, 80
  hull of a set, 95
Measure, 64, 67, 69, 75
  as invariant under isometries, 70
  of sections of a plane set, 76
Medvedev, 59
Menshov, 54, 57, 252, 253, 254
Method of ordering, 107
Metric
  space, 193
  theory of functions, 248
Milman, 255, 257
Molodshy, 124
Mutually complementary in the Young sense, 232
Myshkis, 256

Natanson, 16, 58, 59, 191, 253, 255
Natanson's lemma, 16
Natural order, 107
Nemytsky, 217, 253, 256
Nersesov, 253

Nikolsky, 253
Non-degenerate transformation, 72
Non-vacuousness of the Baire classes, 133
Norm, 195
Normal element of a set, 114
Normed linear space, 195
Novikov, 249, 250, 257
Nowhere dense set, 142
Null class of functions, 128
Number class, 124
  of the first type, 119
  of the second type, 119

Oberfunktion, 158
Open
  circle, 60
  sets, 62
  sphere, 199
Operator, 217
Order type
  of the set of natural numbers, 110
  of the set of rational numbers, 110
Order types, 107
Ordered sets, 107
Ordinal number
  of the first type, 118
  of the second type, 118
Ordinal numbers, 115

Parseval's equality, 55
Perfect set, 61
Perron, 157
  integral, 158, 160
Petrovsky, 209, 251
Pinsker, 257
(P)-integrable function, 159
Plessner, 253, 254, 256, 257
Point
  d-, 57
  fixed (of an operator), 217
  interior, 62
  isolated, 60
  Lebesgue, 16
  limit, 60, 198
  of approximate continuity, 16
  sets in two-dimensional spade, 60
Poisson
  integral, 29
  singular integral, 58
Principle
  Banach's fixed-point, 216
  Cantor-Baire stationary, 145
  of choice, general, 124
Privalov, 22, 28, 253, 254, 257
Projective sets, 250

Rademacher, 27, 253
Radius of a circle, 60
Raikov, 256
Rajchman, 54, 57
Rapprochement operator, 217
Rectifiable curve, 225
Regular contractibility, 103
Regularly ordered subset, 125
Representation of a function by a singular integral, 15
Riemann function, 50
Riemann-Lebesgue theorem, 15
Riemann's
  lemma, 48
  theorem, 51
Riesz, 58
Romanov, 255
Romanovsky, 251, 255
Romanovsky's theorem, 18, 29
Rutman, 255, 257

Saks, 157, 158
Schwarz
  derivatives, 36
  derived numbers, 38
Schwarz's
  lemma, 55
  theorem, 37
Second
  class of functions, 128
  number class, 119
  Schwarz derivative, 37
    derived number, 38

Section of a set, 76, 79
Segment of a set, 111
Selivanovsky, 250
Seliverstov, 253, 254
Semi-continuous functions, 149
Sequence, lacunary, 26
Series
  Fourier, 21, 29
  lacunary trigonometric, 27
  trigonometric, 11, 29
Set
  analytic, 249
  closed, 60
  derived, 60
  diameter of a, 103, 201
  first category, 142
  function, 94
    absolutely continuous, 94
    additive, 94
  measurable hull of a, 95
  of type M, 54, 252
  of type U, 54, 252
  open, 62
  ordered, 107
  perfect, 61
  projective, 250
  segment of a, 111
  well-ordered, 112
Sets, similar, 108
Shilov, 256
Shmulyan, 257
Shneider, 253
Shnirelman, 256
Sierpiński, 124, 127, 245
Similar sets, 108
Similarity mapping, 108
Singular integral, 12
    Bernstein, 58
    de la Vallée Poussin, 58
    Dirichlet, 22
    Fejér, 23
    Kantorovič, 58
    Landau, 58
    Poisson, 58
Smirnov, 253, 257
Solution of a denumberable system of equations, 222
Space
  Banach, 195
  C, 195, 206
  compact, 199
  l_p, 204
  L_p, 212
  m, 196
  s, 197, 202
  S, 198
  strictly normed Banach, 257
Steinhaus, 228
Steinhaus's example, 228
Steklov, 254
  function, 214
Stepanov, 251, 257
Strictly normed Banach space, 257
Subfunction, 158
Subgraph of a function, 83
Sudakov, 216
Sum in the Riemann sense, 51
Summation method
    Abel-Poisson, 28
    Cesàro, 22
    Cesàro-Fejér, 28
Superfunction, 158
Suslin, 249
Symmetric
  derivative, 95
  derived number, 95

Term-by-term integration of a Fourier series, 33
Theorem
  Abel's, 30
  Aleksandrov-Looman, 185, 189
  Arzelà-Ascoli, 207
  Baire's, 143, 146
  Banach's, 217
  Bari-Menshov, 255
  Bari's, 253
  Birnbaum-Orlicz, 232
  Bolzano-Weierstrass, 60
  Borel's, 61

Cantor-Lebesgue, 34
Cantor's, 51, 52, 55
de la Vallée Poussin's, 40
du Bois-Reymond-de la Vallée Poussin, 52
Faddeyev's, 20
Fejér-Lebesgue, 25
Fubini's, 86
Hake's, 178, 185
Hausdorff's, 238
Jordan's, 226
Kolmogorov's, 26, 27, 215
Lebesgue's, 13, 15, 137, 141
Levi's, 88
Luzin-Denjoy, 35
on transfinite induction, 119
Riemann-Lebesgue, 15
Riemann's, 51
Romanovsky's, 18, 29
Schwarz's, 37
Vitali's, 136
Young's, 252
Zermelo's, 125
Tikhonov, 256
Tolstov, 251, 256
Totalization, 157
Transfinite
alephs, 122
induction, 118
theorem on, 119
numbers, 107, 115
Transformation
affine, 72
determinant of a, 72
non-degenerate, 72
Triangle inequality, 193
Trichotomy, 116
Trigonometric series, 11, 29
lacunary, 27
Truncated function, 87
Tulaikov, 212
Type
M set, 54, 252
U set, 54, 252

Uniqueness of trigonometric series expansion, 48
Unit base vectors, 206
Universal function, 139
Unterfunktion, 158
Upper
derivative, 158
semi-continuous function, 150, 152

Value of an operator, 217
Veniaminov, 251
Verchenko, 256
Vilenkin, 254
Vitali covering, 70
Vitali's theorem, 136
Vulikh, 257

Weak convergence, 15
Weierstrass singular integral, 20
Well-ordered sets, 112

Young, 54, 57
Young's
inequality, 231
theorem, 252
Yudin, 257

Zarecki, 251
Zermelo's
axiom, 124, 245
theorem, 125
Zhegalkin, 127, 257
Zygmund, 22, 28, 54

BOOKS ON MATHEMATICS

AKHIEZER, N. I., and GLAZMAN, I. M. *Theory of Linear Operators in Hilbert Space.* Translated by Merlynd Nestell.
Vol. I, $6.25

— Vol. II, *In preparation,* $6.25

BAKER, HENRY F. *Principles of Geometry:* Vol. I: *Foundations.* $5.50

— Vol. II: *Plane Geometry, Conics, Circles, Non-Euclidean Geometry.* $6.50

— Vol. III: *Solid Geometry, Quadrics, Cubic Curves in Space, Cubic Surfaces.* $6.50

— Vol. IV: *Higher Geometry.* Being illustrations of the utility of the consideration of higher space, especially of four and five dimensions. $7.50

— Vol. V: *Analytical Principles of the Theory of Curves.* $6.50

— Vol. VI: *Algebraic Surfaces.* $7.50

DEAUX, Roland. *Introduction to the Geometry of Complex Numbers.* Translated by Howard Eves. $6.50

GOODSTEIN, R. L. *Mathematical Logic.* $3.75

— *Constructive Formalism:* Essays on the Foundations of Mathematics. $3.00

GOODSTEIN, R. L. and PRIMROSE, E. J. F. *Axiomatic Projective Geometry.* $3.25

LIUSTERNIK, L., and SOBOLEV, V. *Elements of Functional Analysis.* Translated by A. E. Labarre, Jr., Herbert Izbicki, and Ward Crowley $8.50

NATANSON, I. P. *Theory of Functions of a Real Variable.* Vol. I (chapters I to IX). Translated and edited by Leo F. Boron and Edwin Hewitt. $8.50

— *Theory of Functions of a Real Variable.* Vol. II (chapters X to XVII with appendices). Translated and edited by Leo F. Boron. $8.50

RIESZ, Frigyes, and SZ.-NAGY, Béla. *Functional Analysis.* Translated by Leo F. Boron. $10.50

— *Appendix to Functional Analysis:* Extensions of Linear Transformations in Hilbert Space Which Extend Beyond This Space, by Béla Sz.-Nagy. $2.50

VAN DER WAERDEN, B. L. *Modern Algebra.* Revised edition.
Vol. I, translated by Fred Blum. $6.00
Vol. II, translated by Theodore J. Benac. $5.50

WAISMANN, Friedrich, *Introduction to Mathematical Thinking.* Translated by Theodore J. Benac. $4.50